Bootstrap for continuous-time autoregressive moving average processes

Von der
Carl-Friedrich-Gauß-Fakultät
der
Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Tobias Niebuhr
geboren am 09. September 1987 in Gifhorn

Eingereicht am: 07.11.2014
Disputation am: 11.12.2014

Referent: Prof. Dr. Jens-Peter Kreiß, TU Braunschweig
Koreferent: Prof. Dr. Efstathios Paparoditis, University of Cyprus

(2014)

Sir, the possibility of successfully navigating an asteroid field is approximately 3,720 to 1.

– C-3PO

Abstract

The main focus of this thesis is to develop bootstrap approaches for the class of continuous-time autoregressive moving average processes.
The first part focuses on the classical setup for time series in discrete time, especially the class of linear processes is considered in detail. Central limit results for integrated periodograms and ratio statistics are shown to change essentially when the estimates are based on low-frequency observations instead of full-time observations. The results correspond to those for continuous-time parameters based on discrete observations. This motivates the further parts of this thesis.
The second part proposes bootstrap possibilities for continuous-time autoregressive processes. Samples of such processes have representations as autoregressions but with uncorrelated innovation sequence only. To circumvent the interdependencies in the innovations, the underlying continuous-time model is used. This allows for a parametric representation with truely independent innovations which is favorable for the bootstrap, however, there is a certain price to pay. Nevertheless, on this basis a valid residual-based bootstrap approach is developed.
The third part considers the bootstrap for continuous-time autoregressive moving average processes. As for pure continuous-time autoregressions, the samples fulfill an autoregessive representation with uncorrelated innovations only. Due to the moving average part, residual-based proposals are not adaptable here. However, the block bootstrap is valid, since it works under very mild assumptions. The additional information on the autoregressive part of the process motivates a two-step bootstrap approach. First, an autoregressive model is fitted. In a second step - to address the remaining dependencies - the block bootstrap is applied. This approach is shown to be valid. Its applicability is tailor-made but not limited to continuous-time autoregressive moving averages. Indeed, it is as widely applicable as standard block bootstraps and thus suitably generalizes the moving block bootstrap. The approach further robustifies the residual bootstrap.

Zusammenfassung

Der Schwerpunkt dieser Arbeit liegt auf der Entwicklung von Bootstrap-Ansätzen für die Klasse der zeitstetigen autoregressiven moving average Prozesse.
Der erste Teil konzentriert sich auf klassische Zeitreihen in diskreter Zeit und dabei insbesondere auf lineare Prozesse. Zentrale Grenzwertresultate für integrierte Periodogramme und Ratio Statistics werden unter der Annahme von Beobachtungen auf niederen Frequenzen entwickelt. Wesentliche Unterschiede zur Situation mit vollständigen Beobachtungen werden aufgezeigt. Die Ergebnisse korrespondieren zu denen für zeitstetige Parameter, die mithilfe von diskreten Beobachtungen geschätzt werden. Diese Eigenschaft motiviert die weiteren Teile dieser Arbeit.
Der zweite Teil stellt Bootstrap-Möglichkeiten für zeitstetige autoregressive Prozesse vor. Beobachtungen solcher Prozesse besitzen eine autoregressive Darstellung, jedoch nur mit unkorrelierten Innovationen. Um die Problematik der Abhängigkeiten innerhalb der Innovationen zu lösen, wird das zugrundeliegende zeitstetige Modell benutzt. Dieses erlaubt eine parametrische Darstellung mit unabhängigen Innovationen, die für Bootstrap-Ansätze günstig ist. Für diese Darstellung ist jedoch ein gewisser Preis zu zahlen. Dennoch kann ein valider Residuen-basierter Bootstrap-Ansatz entwickelt werden.
Der dritte Teil behandelt Bootstrap-Ansätze für zeitstetige autoregressive moving average Prozesse. Wie für stetige autoregressive Prozesse erfüllen solche Stichproben eine autoregressive Darstellung mit nur unkorrelierten Innovationen. Aufgrund des moving average-Anteils lassen sich keine Residuen-basierten Ansätze adaptieren. Die zusätzliche Information über die autoregressive Prozessdarstellung motivert einen zweistufigen Bootstrap-Ansatz. Zunächst wird ein autoregressives Modell angepasst. Als zweites wird der allgemeine Block-Bootstrap angewendet, um die verbliebenen Abhängigkeiten zu adressieren. Die Validität des Ansatzes wird bewiesen. Der Ansatz ist maßgeschneidert, aber nicht beschränkt auf die zeitstetigen autoregressiven moving average Prozesse. Genauer ist die Methode genauso allgemein anwendbar wie der Standard-Block-Bootstrap. Der Ansatz erweitert den Moving-Block-Bootstrap sinnvoll und robustifiziert außerdem den Residuen-basierten Bootstrap.

Contents

List of Figures

1 Introduction

For hundreds of years, people have been recording data of their daily life. The examples of observed data are highly numerous and cover wide ranges – from high water levels over unemployment rates to stock prices and far beyond. Of course, data recording needs expenditure and the question of the use of such recordings arises immediately. One main reason certainly is the will to extract meaningful information of the afore-recorded observations, like e.g. the prediction of tomorrow's wheather conditions. Consequently, when wanting to get to know information about future events, techniques to describe and analyze the collected data are needed. This is where mathematical time series analysis comes into play.

Classical time series analysis assumes a certain number of observations $X_1, \dots, X_n$ stemming from an underlying stochastic process $X = \{X_t : t \in \mathbb{Z}\}$ on some probability space $(\Omega, \mathcal{F}, P)$. To guarantee a certain amount of ability to analyze interdependencies of such data, which involves finding appropriate models for the dependence structure of the process X at hand, classical time series analysis assumes the stochastic process to fulfill some structural properties. The probably most popular assumption is the one on stationarity.

The autocovariance function is the most convenient way to describe the inner dependencies of time series. Using the autocovariance the concept of stationarity can be defined.

Definition 1.1. *Let $X = \{X_t : t \in \mathbb{Z}\}$ be a real-valued time series such that $Var(X_t) < \infty$ for each $t \in \mathbb{Z}$, then the autocovariance function $\gamma_X(\cdot, \cdot)$ of X is defined by*

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \quad r, s \in \mathbb{Z}. \tag{1.1}$$

Definition 1.2. *The real-valued time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be a stationary time series if*

(i) $E|X_t|^2 < \infty$ *for all* $t \in \mathbb{Z}$,

(ii) $EX_t = \mu$ *for all* $t \in \mathbb{Z}$,

(iii) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ *for all* $r, s, t \in \mathbb{Z}$.

In cases of stationary time series it thus suffices to denote the autocovariance in dependence on the so-called lag instead of two specific time points. Therefore for stationary time series we will use the notation $\gamma_X(r - s)$ instead of $\gamma_X(r, s)$ in the following.

We have defined stationarity only for realizations of the time series at integer time points. It is not difficult to transfer the idea of stationarity to more general index sets, as e.g. high-frequency index sets.

In the literature, stationarity, as just defined, often is referred to as weak stationarity. Without further qualification, the term stationarity will always refer to the definition above throughout this thesis. As one might expect after having defined weak stationarity there is another even more restrictive definition on strong stationarity.

Definition 1.3. *The real-valued time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, \ldots, X_{t_k})$ and $(X_{t_1+h}, \ldots, X_{t_k+h})$ are the same for all positive integers k and for all $t_1, \ldots, t_k, h \in \mathbb{Z}$.*

Strict stationarity together with finite second moments of time series directly gives weak stationarity. Roughly speaking, while (weak) stationarity requires time series to fulfill equal second-order properties independently on the time indices, strict stationarity ensures that any characteristics of the time series – even the ones beyond second-order properties – are competetive from one time interval to another.

During the latest decades the importance of time series analysis has been growing. A huge amount of data is recorded every second and the question for techniques to extract the information of interest is more present than ever. Extensively increasing computing powers allow for finer and more feasible models. Since the turn of the millennium especially continuous-time modeling has become of growing interest. In

this thesis, we present bootstrap possibilities for the wide class of continuous-time autoregressive moving average processes. However, to find a common starting point a short overview over important results and classes of discrete time series is given. For more detailed expositions on these topics see Brockwell and Davis (1991) or more recent Kreiss and Neuhaus (2006). After a short review on time series analysis and on the bootstrap, the main concepts of this thesis are outlined.

1.1 Preliminary definitions

This section will shortly review well-established results in time series analysis. We will restrict ourselves to univariate real-valued time series for notational reasons. However, all results presented can easily be adapted to the multivariate pattern and/or complex-valued time series.

When considering time series one usually associates an underlying model which is distorted by some random disturbance parameter. This disturbance is usually called a white noise sequence.

Definition 1.4. *The time series* $X = \{X_t : t \in \mathbb{Z}\}$ *is said to be a white noise if* $EX_t = 0$ *and* $EX_t^2 = \sigma^2 < \infty$ *for all* $t \in \mathbb{Z}$ *and* $Cov(X_t, X_s) = 0$ *for all* $t \neq s \in \mathbb{Z}$.

By definition a white noise is weakly stationary but not necessarily strictly stationary. When further assuming the white noise to be an independent and identically distributed (i.i.d.) sequence, the conditions for strict stationarity are fulfilled as well.

While the name *noise* is more or less obvious for e.g. the context of signal processing, the characteristic *white* is less straightforward to understand. So far, we only considered time-domain methods. Another way to investigate time series is the frequency-domain. The most important representant in the frequency-domain may be the spectral density. This quantity is closely related to the autocovariance function. Let the autocovariance be absolutely summable, i.e. $\sum_{h\in\mathbb{Z}} |\gamma(h)| < \infty$, then the spectral density $f(\omega)$ can be represented by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}, \quad \omega \in (-\pi, \pi]. \tag{1.2}$$

The spectral density gives advice about preferred and supressed frequencies of the time series. For the white noise, the spectral density in (1.2) reduces to a single

summand, namely the one for $h = 0$, and thus results in some constant on $(-\pi, \pi]$. This can be interpreted as a uniform spectrum with no preferred frequency or wavelength. This is why time series as in Definition 1.4 are called white noise.

Some famous time series models are in order now. Autoregressive models are charming in application cases since they are easily to interpret. They weight previous realizations of the time series and add an additional random disturbance. In contrast, moving average models are suitable to model the occurence of shocks and cases of finite dependence. The formal models are as follows.

Definition 1.5. *The time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be a linear process if it exhibits the representation*

$$X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \tag{1.3}$$

where e_t is an i.i.d. white noise sequence, $b_0 = 1$ and the sequence $(b_j)_{j\in\mathbb{Z}}$ are summable model coefficients. If $b_j = 0$ for all $j < 0$, the time series is called causal.

Definition 1.6. *The time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be an autoregressive process of order p (AR(p)) if it yields the representation*

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + e_t, \quad \forall t \in \mathbb{Z}, \tag{1.4}$$

where e_t is an i.i.d. white noise sequence and $a_1, \ldots, a_p$ are the model coefficients.

Definition 1.7. *The time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be a moving average process of order q (MA(q)) if it yields the representation*

$$X_t = \sum_{j=0}^{q} b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \tag{1.5}$$

where e_t is an i.i.d. white noise sequence, $b_0 = 1$ and $b_1, \ldots, b_q$ are the model coefficients.

Definition 1.8. *The time series $X = \{X_t : t \in \mathbb{Z}\}$ is said to be an autoregressive moving average process of orders (p, q) (ARMA(p,q)) if it yields the representation*

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=0}^{q} b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \tag{1.6}$$

where e_t is an i.i.d. white noise sequence, $b_0 = 1$ and $a_1, \ldots, a_p, b_1, \ldots, b_q$ are the model coefficients.

Setting the model orders p and/or q to infinity is allowed as well. Under suitable conditions on the coefficients of the AR(p)-model and the ARMA(p,q)-model they yield stationarity. Further, for suitable coefficients they have representations as causal linear processes. For computational reasons, this ability is very useful and will often be used within this thesis.

1.2 Bootstrap

The analysis of samples stemming from a time series always comes along with the question about how to handle uncertainty and randomness. Consistent estimates for specific statistics might be a first step to circumvent random effects. However, a single estimate might not be sufficient for reliable statements on the statistics behavior, such as e.g. confidence intervals. This could be extracted if one knew the underlying distribution. Consequently, one main task lies in suitably approximating the unknown distributions of statistics of interest. In several cases, the limiting behavior of statistics is known or can be obtained analytically from central limit theorems. The straightforward approach is then to estimate the parameters of the limiting distribution conditional on the given sample at hand. However, when estimating quantities the question about how close these estimates are to the true unknown values arises immediately. Nevertheless, its ease of use makes this approach somehow charming but approximation errors might be essentially large.

Considering this scenario in more detail, many statistics yield limiting normal distributions but estimates based on finite sample sizes might crucially differ from normal distributions and keep high skewness. Then, the ordinary estimate of the limiting normal distribution, the so-called normal approximation, forces the approximation to two unsatisfying errors. First, the approximation error stemming from estimates based on the finite sample arises, secondly, the normal approximation ensures a symmetric approximated distribution which cannot address for possibly occuring skewness of the true finite sample distribution.

Rapid improvements in technologies and especially reduction of computational time allow for new approaches which could not be thought of fifty years ago. During the last decades, so-called subsampling procedures have become widely accepted tools for the estimation of statistics and the approximation of their distributions. Motivated by the aforementioned, Efron (1979) introduced the so-called *bootstrap* to

circumvent the inaccurancies of the normal approximation and further to improve the reliability of estimates. This seminal work stated a new point of view. Its main idea is as follows.

Assume we are given a sample of i.i.d. random variables $X_1, \ldots, X_n$ and the interest is on the distribution of the statistic $T_n = T_n(X_1, \ldots, X_n)$. Then, sample with replacement n times from the original data set and obtain a bootstrap data set $X_1^*, \ldots, X_n^*$. On this basis, compute the bootstrap statistic $T_n^* = T_n(X_1^*, \ldots, T_n^*)$. Repeating these steps very frequently determines the empirical distribution of T_n^*, which is a suitable estimator for the true distribution of T_n^*. To justify this proposal, one analytically proofs that the distributions of T_n^* and of T_n are close together. In practice, this is often shown by computing the central limit results for T_n and T_n^*. If both yield the same limiting distribution, the estimates will coincide asymptotically. Then we say that the *bootstrap works*. However, this does not show that the bootstrap is in advantage over the normal approximation, although very many simulation studies have shown that bootstrap proposals perform very well in comparison to normal approximations. Analytical validation of bootstrap advantages is rather difficult.

One big advantage of the bootstrap is that it approximates the finite sample distribution and not the limiting distribution. Thus, it is able to address skewness of distributions very well. Further, its application is directly settled on the sample itself resulting in an easy way of use.

Unfortunately, Efron's procedure is restricted to i.i.d. data, which means that in the context of dependent data, as e.g. time series, it is not directly applicable. Proofs for bootstrapping time series data are neither trivial nor straightforward. Nevertheless, general approaches for bootstrapping dependent data were proven. We give a short overview about some well-established bootstrap procedures for time series. Each approach has its own advantages and disadvantages, which will tip the balance to prefer either the one or another in different scenarios.

- **Residual bootstrap.** The main idea of the residual bootstrap is to fit a paramteric model to the data and to apply Efron's i.i.d. bootstrap to the obtained estimated residuals. Consider e.g. an AR time series and an AR model fit. The residual bootstrap aims at pinpointing the underlying parametric model and resampling the i.i.d. innovations. These approaches are tailor-made

for situations in which the underlying time series model is known. It usually delivers remarkably good performances in simulation studies. Otherwise, this approach has to be considered carefully, since it is strictly limited to the process class considered. Even misspecification of the model order might lead to approximation mistakes. For references see Davis (1977), Freedman (1984), Efron and Tibshirani (1986), Bose (1988), Kreiss (1988), Kreiss (1997), Bühlmann (1997), Kreiss, Paparoditis and Politis (2011), or the references therein.

- **Block bootstrap.** The main idea of the block bootstrap is very similar to the i.i.d. bootstrap. The sample of length n is cut into blocks of length l. Then, one resamples (with replacement) the blocks of the time series rowing them up to obtain a bootstrap sample. The motivation for this approach is two-sided. On the one hand, the interdependencies within the blocks are assured to hold. On the other hand, the resampling of the blocks ensures a sufficient degree of randomness. These two competing components of the procedure have to be balanced by suitable choices for the block length l and the corresponding number of blocks b. Usually, they are both required tending to infinity with increasing sample size. For the special case $l = 1$, the block bootstrap reduces to the i.i.d. bootstrap. So far, the block bootstrap is possibly the most general bootstrap procedure and only requires very mild assumptions on the underlying process. However, if parametric approaches are possible, the block bootstrap usually performs less accurate in comparative simulation studies. Many studies investigating and modifying different aspects of the block bootstrap can be found. For references see Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992), Shao and Yu (1993), Naik-Nimbalkar and Rajarshi (1994), Bühlmann (1994), Bühlmann and Künsch (1995), among others.

- **Frequency bootstrap.** In comparison to the residual and block bootstrap procedures, which both are time-domain approaches, the frequency-based bootstrap procedures rely on the asymptotic features of the periodogram. It is widely known that the periodogram realizations evaluated at different frequencies are asymptotically independent and thus approaches settled on these realizations do not require any parametric assumption. These approaches are shown to perform reasonably in simulations but are somehow limited to functions of the periodogram since they are not able to produce bootstrap replicates in the time domain. For references see Franke and Härdle (1992), Dahlhaus

and Janas (1996), Kreiss and Paparoditis (2003), Shao and Wu (2007), and references therein.

- **Other bootstrap procedures.** In addition to the aforementioned approaches several other bootstrap methods have been stated, each of them introducing its own key aspect, as e.g. modification of the resampling mechanism for the innovations or switchings between the frequency and the time domain. Among them are the wild bootstrap of Wu (1986), the dependent wild bootstrap of Shao (2010) and the TFT-bootstrap of Kirch and Politis (2011), along with others. If we require any of these approaches, they will be discussed later on.

1.3 Main results of this thesis

The contribution of this thesis is in several aspects. To start with, in chapter 2 we will investigate linear processes which are observable on a lower frequency only. At first glance, one might be surprised why discrete-time processes are considered in a thesis about a continuous-time model. However, there are relevant parallels. It will be shown that estimates based on low-frequency observed samples are neither able to describe the process characteristics in between the observations nor to describe the interdependencies between the observed process and time points in between. Crucial information about the underlying process' structure is missing and furthermore asymptotics for estimates will vary. As a consequence, estimates based on low-frequency samples will highly depend on the observation rhythm. Furthermore, bootstrap approaches and the identification of the underlying model are no longer straightforward and require strenuous efforts. Chapter 2 gives central limit results based on low-frequency samples for the general class of integrated periodograms. Under the assumption of an underlying AR process bootstrap proposals will be discussed.

The investigation of low-frequency samples naturally motivates the main parts of this thesis. While chapter 2 settles on a given process which is observed on a larger time grid, the remainder of this thesis considers this scenario from another point of view. The motivation is to have a given sample at hand and to assume that there are observations in between the sampled time points. Using a fill-in idea one directly is lead to assume the unknown underlying process as continuous.

Chapter 3 introduces the continuous-time autoregressive moving average model. Besides the general model, main results on this class of processes and especially on representations of discrete samples are shortly reviewed.

Chapter 4 restricts its investigation to continuous-time autoregressions instead of continuous-time autoregressive moving averages. This class of processes has been of interest to physicists and engineers for many years (see e.g. Fowler (1936)). The neglection of the moving average coefficients essentially eases the model. It will be shown that observing continuous-time autoregressions has many parallels to low-frequency sampling in discrete time. Furthermore, the problem corresponds to multivariate processes from which only one component is observable. A valid bootstrap procedure based on samples stemming from continuous-time autoregressions will be presented for the general autocovariance and autocorrelation.

Chapter 5 introduces the autoregressive-aided block bootstrap. This method is motivated by representations of samples of continuous-time autoregressive moving average processes. For such samples as well as for low-frequency sampled discrete-time autoregressions, the bootstrap proposal is tailor-made. However, its applicability goes far beyond. Methods from both the residual bootstrap and the block bootstrap will be combined resulting in a more general approach from which both procedures can be obtained as special cases. This bootstrap approach will be shown to be valid for the large class of functions of generalized means as used by Künsch (1989). Although the main idea is to improve the block bootstrap, the approach can also be shown to robustify the residual bootstrap.

2 | Low-frequency observations of linear processes

Based on : Niebuhr, T. and Kreiß, J.-P. (2014):
Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies.
The International Statistical Review, Vol. 82, No. 1, pp. 123-140.

Abstract. One of the most frequently used class of processes in time series analysis is the one of linear processes. For many statistical quantities, among them sample autocovariances and sample autocorrelations, central limit theorems are available in the literature. We investigate classical linear processes under a non standard observation pattern, namely we assume that we are only able to observe the linear process at a lower frequency, i.e. only every K-th data point is observable for some $K \in \mathbb{N}$. It is shown that such observation pattern destroys the linear structure of the observations and leads to substantially different asymptotic results for standard statistical quantities. Central limit theorems are given for sample autocovariances and sample autocorrelations as well as more general integrated periodograms and ratio statistics. Moreover for specific autoregressive processes the possibilities to estimate the parameters of the underlying autoregression from lower frequency observations are addressed. Finally we suggest for autoregressions of order two a valid bootstrap procedure. A small simulation study demonstrates the performance of the bootstrap proposal for finite sample size.

2.1 Introduction

The estimation of autocorrelation plays an important role in time series analysis. Bartlett (1946) already derived an explicit formula for the asymptotic variance-covariance matrix of sample autocorrelations for linear processes. In this part of the present thesis, as a linear process we denote a time series $(X_t : t \in \mathbb{Z})$ which allows for a (possibly infinite order) moving average representation with independent and identically distributed (i.i.d.) noise process (cf. (2.1) for a precise definition). This assumption is in contrast to the classical assumption of *white noise* innovations, which are uncorrelated, only. The result of Bartlett (1946), which contains as an important contribution (known as *Bartlett's formula*) that for linear processes the asymptotic variance-covariance matrix of sample autocorrelations only depends on the second order structure of the underlying stationary time series (that means only on the autocorrelation function), is taught in nearly every introductory course to time series analysis. For an exemplary textbook reference we refer to Brockwell and Davis (1991), Theorem 7.2.1.

Romano and Thombs (1996) pointed out that the classical Bartlett's formula may go completely wrong for series containing any form of nonlinearity. However, the classical Wold decomposition (cf. Brockwell and Davis (1991), Section 5.7) states that every purely non-deterministic (weakly) stationary time series admit an infinite order moving average representation with square summable coefficients and with white noise process, i.e. with centered and uncorrelated noise sequence. It is worth mentioning that the white noise sequence in the Wold decomposition generally consists by no means of independent random variables. So the class of linear processes is considerably smaller than the class of purely non-deterministic stationary processes. Recent work of Francq and Zakoïan (2009) investigates the asymptotic covariance structure of sample autocovariances and autocorrelations in case that the i.i.d. structure for the white noise in a moving average representation of the underlying time series is not fulfilled. The authors extended Bartlett's formula by giving the asymptotic covariances between sample autocovariances as well as between sample autocorrelations in function of characteristics of the underlying process under quite mild regularity assumptions. This result substantiates the differences to the classical linear process case. Especially their Theorem 2 shows that in general the theoretical autocorrelation structure of the underlying process is not sufficient to characterize the asymptotic covariances of sample autocorrelations, as it is the case for classical

linear processes. As another example Dunsmuir and Robinson (1981a) and (1981b) considered classical linear time series with randomly missing observations and discussed the influence to the asymptotic behaviour of sample autocovariances and autocorrelations. Kokoszka and Politis (2011) considered autoregressive processes with conditional heteroscedasticity (ARCH) and stochastic volatility (SV) models. They showed that Bartlett's formula is not warranted for such models.

In this part of the present thesis we start from classical linear processes, but we approach the problem of possibly occurring non-linearity from another point of view. It will be assumed that we are able to observe the underlying linear process at a lower frequency, only. This means that we for example are only able to observe every K^{th} data point, $K \in \mathbb{N}$; thus, our observations consist of $Y_t = X_{K\,t}$, $t = 1, \ldots, n$. As we will see in the following, the asymptotic behavior of sample autocovariances and sample autocorrelations changes in our setup and especially we loose the convenient Bartlett formula for sample autocorrelations. This result immediately implies that lower frequency observations from linear processes typically lie not again in the class of linear time series. It is worth mentioning that a similar phenomenon occurs when dealing with discrete observations from continuous time moving average processes (cf. Cohen and Lindner (2013)).

From an applied point of view our lower frequency observation scheme could mean that we can (or want to) observe a daily time series only every second (or generally every K^{th}) day. From another point of view our assumption models a situation in which we have time series data $Y_1, \ldots, Y_n$ (think of weekly data) at hand, but for which the assumption of linearity is only fulfilled at a higher frequency (think of daily or hourly data). Then our observations at hand exactly are realizations from a linear process observed at a lower frequency. Thinking again of applications in which we typically can select to a certain extent the frequency (hourly, daily, weekly, monthly and so on) at which we can observe our data, it should be viewed as a fortunate instance that the assumption of linearity just exactly holds for the time frequency we have chosen or which is given for our observations. In most cases one has to face the situation that the underlying observations at most are realizations from a linear process observed at a lower frequency.

Besides classical sample autocovariances and sample autocorrelations we consider in detail the class of integrated periodograms and ratio statistics introduced by

Dahlhaus (1985). It is obtained that we observe similar phenomenons for such statistics. Furthermore we investigate for a special class of linear autoregressive processes the opportunities to estimate from observations at lower frequencies the autoregressive parameters of the underlying process and the abilities of the bootstrap for such situations.

The rest of the chapter is organized as follows. Section 2.2 briefly reviews the relevant classical results for ordinary linear processes. In Section 2.3 we derive the corresponding results for linear processes under the assumption of observations at a lower frequency. A special autoregressive situation is considered in Section 2.4, where we also discuss estimation of parameters of the underlying (and not completely observable) process as well as a specific bootstrap approach. The results of a small simulation study are reported in Section 2.4.2. All technical proofs are deferred to the proof Section 2.6.

2.2 Review of results for classical linear processes

As already discussed in the introduction of this chapter the class of linear processes is one of the most popular classes in time series analysis. A *(strong) linear process* is defined as follows.

Definition 2.1. *A real valued stochastic process $X = (X_t : t \in \mathbb{Z})$ in discrete time possessing the representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} b_j e_{t-j}\,,\ t \in \mathbb{Z}\,, \tag{2.1}$$

with $\mu \in \mathbb{R}$ and absolutely summable coefficients $b_j \in \mathbb{R}$, that is $\sum_{j=-\infty}^{\infty} |b_j| < \infty$, and with an i.i.d. process $(e_t : t \in \mathbb{Z})$, with $E\, e_t = 0$ and $E\, e_t^2 = \sigma_e^2 \in (0, \infty)$, is called a strong linear process. If $b_j = 0$ for all $j < 0$, then the time-invariant filter $(b_j : j \in \mathbb{Z})$ is called causal.

Remark 2.2. *This definition is slightly different to Definition 1.5 and assumes for absolute summability of the coefficients. Anyhow, for the remainder of this chapter we suggest by a linear process a process as given in Definition 2.1.*

For linear processes the autocovariance function reads

$$\gamma_X(h) = \sigma_e^2 \sum_{j=-\infty}^{\infty} b_j b_{j+h}\,. \tag{2.2}$$

The sample autocovariance based on observations $X_1, \ldots, X_n$ is

$$\widehat{\gamma}_X(h) \;=\; \frac{1}{n}\sum_{t=1}^{n-|h|}\Big(X_{t+|h|}-\overline{X}\Big)\Big(X_t-\overline{X}\Big) \tag{2.3}$$

and the sample autocorrelation will be denoted by

$$\widehat{\varrho}(h)=\frac{\widehat{\gamma}_X(h)}{\widehat{\gamma}_X(0)},\quad h\in\mathbb{N}, \tag{2.4}$$

as long as $\widehat{\gamma}_X(0)>0$. The following result is then well-known.

Theorem 2.3. *Let (X_t) be a linear process and assume that $Ee_t^4 := \eta\,\sigma_e^4 < \infty$. Then, for $h,k\in\mathbb{N}_0$ and as $n\to\infty$:*

(i) It holds

$$E\,\widehat{\gamma}_X(h)=\gamma_X(h)+\mathcal{O}\left(n^{-1}\right) \tag{2.5}$$

(ii) Furthermore

$$n\cdot Cov(\widehat{\gamma}_X(h),\widehat{\gamma}_X(k))\to c_X(h,k)\,, \tag{2.6}$$

where

$$\begin{aligned} c_X(h,k) \;&=\; (\eta-3)\gamma_X(h)\gamma_X(k) \\ &\quad+\sum_{r=-\infty}^{\infty}\Big(\gamma_X(r)\gamma_X(r+k-h)+\gamma_X(r+k)\gamma_X(r-h)\Big)\,. \end{aligned} \tag{2.7}$$

(iii) Moreover for $M\in\mathbb{N}$

$$\sqrt{n}\Big(\widehat{\gamma}_X(h)-\gamma_X(h)\;:\;h=0,...,M\Big)\xrightarrow{\mathcal{D}}\mathcal{N}(0,S_M)\,, \tag{2.8}$$

where $\xrightarrow{\mathcal{D}}$ denotes weak convergence and

$$S_M=\Big(c_X(h,k)\Big)_{h,k=0,\ldots,M} \tag{2.9}$$

with $c_X(h,k)$ as in (2.7).

The theorem can be found as Proposition 7.3.1 and Proposition 7.3.4 in Brockwell and Davis (1991). Theorem 2.3 can be extended to so called integrated periodograms

or spectral means. Recall that based on observations $X_1, \ldots, X_n$ the periodogram $I_n(\omega)$ is defined by

$$I_n^X(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-i\omega t} \right|^2 , \ \omega \in [0, \pi] . \tag{2.10}$$

Then integrated periodogram estimators are given by

$$M(I_n^X, W) = \int_0^\pi W(\omega) I_n^X(\omega) d\omega , \tag{2.11}$$

where W denotes an appropriately defined function on $[0, \pi]$. Such integrated periodograms and also so called ratio statistics

$$R(I_n^X, W) = \frac{M(I_n^X, W)}{M(I_n^X, 1)} , \tag{2.12}$$

which are normalized spectral means, have been defined and investigated in full detail in Dahlhaus (1985). For linear processes Dahlhaus proved the following asymptotic properties of (2.11).

Theorem 2.4. *Let (X_t) be a linear process and assume that $Ee_t^4 := \eta \sigma_e^4 < \infty$. Denote by f_X the spectral density of the process (X_t) and assume $\sum_{j=-\infty}^{\infty} |j||b_j| < \infty$. The function $W : [0, \pi] \to \mathbb{R}$ is assumed to be of bounded variation. Then*

$$\sqrt{n}\Big(M(I_n^X, W) - M(f_X, W) \Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2) , \tag{2.13}$$

where

$$\tau^2 = (\eta - 3) \left(\int_0^\pi W(\omega) f_X(\omega) \, d\omega \right)^2 + 2\pi \int_0^\pi W^2(\omega) f_X^2(\omega) \, d\omega . \tag{2.14}$$

Remark 2.5. *Note that if $W(\omega) = 2\cos(\omega h)$, $h \in \mathbb{N}_0$, then (2.11) is the sample autocovariance $\widehat{\gamma}_X(h)$ and Theorem 2.4 coincides with the results of Theorem 2.3. The class of integrated periodograms includes many other statistics quite common in time series analysis among them spectral distributions and smoothed periodograms. It should be noted that the simple sample mean $\overline{X}$ cannot be represented as an integrated periodogram.*

It is emphasized that the most important advantage of the transition from integrated periodograms to ratio statistics is the fact that the fourth order cumulant of the i.i.d. noise process η does not show up in the limiting variance of ratio statistics. Thus the asymptotic distribution of ratio statistics does not depend on any characteristic of the underlying linear process beyond second order properties. The precise result is as follows (cf. Dahlhaus and Janas (1996) or Kreiss and Neuhaus (2006), Corollary 10.17).

Corollary 2.6. *Under the assumptions of Theorem 2.4 we have that*

$$\sqrt{n}\left(R(I_n^X, W) - R(f_X, W)\right) \tag{2.15}$$

has a limiting Gaussian distribution with zero mean and variance given by

$$\widetilde{\tau}^2 = 2\pi \int_0^\pi \widetilde{W}^2(\omega) f_X^2(\omega)\, d\omega\,, \tag{2.16}$$

where

$$\widetilde{W}(\omega) = \frac{W(\omega)\int_0^\pi f_X(\lambda)\, d\lambda - \int_0^\pi W(\lambda) f_X(\lambda)\, d\lambda}{\left(\int_0^\pi f_X(\lambda)\, d\lambda\right)^2}\ ,\ \omega \in [0,\pi]\,. \tag{2.17}$$

Remark 2.7. *It should be noted that Theorem 2.4 as well as Corollary 2.6 are valid under the assumption of a linear process, only.*

2.3 Linear processes observed at lower frequencies

This section will generalize the results of Section 2.2. For reasons of simplicity and easier notation throughout the whole section we assume the linear process (X_t) having a causal time-invariant filter and thus a one-sided moving average representation (cf. Definition 2.1). In contrast to the usual and well-studied situation with observations at times $\{1, 2, \ldots, n\}$, we now assume to have observations at times $\{1K, 2K, \ldots, nK\}$ at hand (compare with the assumptions in Politis (1993)). The resulting sample $Y_1, \ldots, Y_n$ is the partial process $Y_t = X_{Kt}$ of the underlying linear process (X_t) sampled at a lower frequency. The sampled process possesses the representation

$$Y_t = X_{Kt} = \sum_{i=0}^{K-1} \sum_{j=0}^{\infty} b_{Kj+i} e_{K(t-j)+i} =: \sum_{i=0}^{K-1} Y_t^{(i)}. \tag{2.18}$$

It is worth mentioning that the processes $(Y_t^{(i)})$ and $(Y_t^{(j)})$ are mutually independent if $i \neq j$. Moreover, each $Y_t^{(i)} = \sum_{j=0}^{\infty} b_{Kj+i} e_{K(t-j)+i}$ indeed is a linear process. Thus, the observed process (Y_t) can be represented as a sum of K stochastic independent linear processes.

However, (Y_t) itself is not a strong linear process in general. To obtain examples where (Y_t) is not strongly linear, one can use the elementary arguments used in Section 3.2 of Francq and Zakoïan (1998). One exception where (Y_t) itself is linear is

given if the underlying linear process (X_t) is Gaussian. Another exception arises if (X_t) is assumed to be a moving average process of order less than K. The resulting partial process (Y_t) is independent and thus is a strong white noise.

Example 2.8. *Let (X_t) be a linear process and consider the simplest case $K = 2$. We obtain for the observed process (Y_t) the representation*

$$Y_t = b_0\, e_{2t} + b_1\, e_{2t-1} + b_2\, e_{2t-2} + b_3\, e_{2t-3} + \dots\,, \tag{2.19}$$

which coincides with the linear filter $(b_0, b_1, b_2, \dots)$ applied to the innovation sequence $(e_{2t}, e_{2t-1}, e_{2t-2}, \dots)$. If (Y_t) were linear then it would follow that Y_{t+1} is the same filter applied to the shifted sequence $(e_{2t+1}, e_{2t}, e_{2t-1}, \dots)$, but Y_{t+1} in fact can only be written as the filter applied to the double shifted sequence $(e_{2t+2}, e_{2t+1}, e_{2t}, \dots)$. Thus, (Y_t) itself is not a linear process.

Considering the K independent processes $(Y_t^{(i)})$, $i = 0, \dots, K-1$, and their individual autocovariances

$$\gamma_{Y^{(i)}}(h) \equiv \gamma_Y^{(i)}(h) = \sigma_e^2 \sum_{j=0}^{\infty} b_{Kj+i} b_{Kj+i+Kh} \tag{2.20}$$

we obtain that the autocovariances of (Y_t) are given by

$$\gamma_Y(h) = \sum_{i=0}^{K-1} \gamma_Y^{(i)}(h) = \sigma_e^2 \sum_{j=0}^{\infty} b_j b_{j+Kh} = \gamma_X(Kh). \tag{2.21}$$

Denote the sample autocovariance of Y_t as

$$\widehat{\gamma}_Y(h) \;=\; \frac{1}{n} \sum_{t=1}^{n-|h|} \left(Y_{t+|h|} - \overline{Y}\right)\left(Y_t - \overline{Y}\right). \tag{2.22}$$

It should be noted that $\widehat{\gamma}_Y(h)$ as given in (2.22) does not coincide with the corresponding sample autocovariances $\widehat{\gamma}_X(Kh)$ (based on a complete sample $X_1, \dots, X_n$). The following lemma specifies the asymptotic covariance matrix of the distribution of $\widehat{\gamma}_Y(h)$. In general, it arises from the non-linearity of (Y_t) that it differs from (2.7).

Lemma 2.9. *Let (X_t) be a linear process (cf. Definition 2.1) with $Ee_t^4 := \eta\sigma_e^4 < \infty$. Further, let the observations $Y_t = X_{Kt},\ t = 1, \dots, n,\ (K \in \mathbb{N}$ fixed) be given. Then for $h, k \in \mathbb{N}_0$ and as $n \to \infty$:*

$$E\,\widehat{\gamma}_Y(h) = \gamma_Y(h) + \mathcal{O}\left(n^{-1}\right), \tag{2.23}$$

and

$$n \cdot Cov(\widehat{\gamma}_Y(h), \widehat{\gamma}_Y(k)) \to c_Y(h,k), \tag{2.24}$$

where

$$\begin{aligned} c_Y(h,k) &= (\eta-3)\sum_{i=0}^{K-1}\left(\gamma_Y^{(i)}(h)\gamma_Y^{(i)}(k)\right) \\ &\quad + \sum_{r=-\infty}^{\infty}\left(\gamma_Y(r)\gamma_Y(r+k-h)+\gamma_Y(r+k)\gamma_Y(r-h)\right). \end{aligned} \tag{2.25}$$

Note that (2.25) differs from the corresponding expression (2.7) for linear processes by the first summand only. However this difference is rather relevant as we will see later on.

Using similar techniques as for the linear process case (cf. Brockwell and Davis (1991), proofs of Theorem 7.2.1 and Proposition 7.3.3) we easily are able to establish asymptotic normality for sample autocovariances $\widehat{\gamma}_Y(h)$.

Theorem 2.10. *Let (X_t) be a linear process and assume that $Ee_t^4 := \eta\,\sigma_e^4 < \infty$. For observations $Y_1, \ldots, Y_n$ as in Lemma 2.9 we have for every $M \in \mathbb{N}_0$ and as $n \to \infty$ that $\sqrt{n}(\widehat{\gamma}_Y(h) - \gamma_Y(h))_{h=0,\ldots,M}$ converges to a multivariate Gaussian distribution with mean zero and covariance matrix given by*

$$S_M = (c_Y(h,k))_{h,k=0,\ldots,M}\,, \tag{2.26}$$

where $c_Y(h,k)$ is given in (2.25).

Remark 2.11. *We already mentioned that the asymptotic covariance matrices for sample autocovariances of the linear process (X_t) and the process $(Y_t) = (X_{K\,t})$ substantially differ. For the special case $K = 2$, this means that we can observe the linear process X_t at even time points, only. A comparison of the two asymptotic variances of $\widehat{\gamma}_Y(h)$ and $\widehat{\gamma}_X(2h)$, cf. (2.25) and (2.7), respectively, yields*

$$\begin{aligned} c_X(2h,2h) &= c_Y(h,h) \\ &\quad +(\eta-3)\,\sigma_e^4\left(\sum_\nu b_{2\nu+1}b_{2\nu+1+2h}\sum_r b_{2r}b_{2r+2h}\right)^2 \\ &\quad +\sum_r\left(\gamma_X^2(2r+1)+\gamma_X(2(r+h)+1)\gamma_X(2(r-h)+1)\right). \end{aligned} \tag{2.27}$$

Note that lag h in terms of Y_t corresponds to lag $2h$ in terms of X_t. It appears that for the partial (even, say) process Y_t the asymptotic variance of sample autocovariances only consists of even terms so that the difference to the asymptotic variance of sample autocovariances of the entire process X_t just consists of the odd parts.

Remark 2.12. *Another representation equivalent to (2.25) and now for the general case $K \in \mathbb{N}$ reads*

$$\begin{aligned} c_Y(h,h) &= (\eta-3)\,\sigma_e^4 \sum_{i=0}^{K-1} \left(\sum_{j=-\infty}^{\infty} b_{Kj+i} b_{K(j+h)+i} \right)^2 \qquad (2.28) \\ &\quad + \sum_{r=-\infty}^{\infty} \Big(\gamma_Y^2(r) + \gamma_Y(r+h)\gamma_Y(r-h) \Big) \\ &= (\eta-3)\,\sigma_e^4 \sum_{i=0}^{K-1} \left(\sum_{j=-\infty}^{\infty} b_{Kj+i} b_{K(j+h)+i} \right)^2 \qquad (2.29) \\ &\quad + \sum_{r=-\infty}^{\infty} \Big(\gamma_X^2(Kr) + \gamma_X(K(r+h))\gamma_X(K(r-h)) \Big). \end{aligned}$$

It is worth mentioning that the first summand (2.29) cannot be represented in terms of $\gamma_X(\cdot)$. In contrast to (2.28) we have

$$\begin{aligned} & c_X(Kh,Kh) \qquad (2.30) \\ &= (\eta-3)\,\gamma_X^2(Kh) + \sum_{r=-\infty}^{\infty} \Big(\gamma_X^2(Kr) + \gamma_X(K(r+h))\gamma_X(K(r-h)) \Big) \\ &= (\eta-3)\,\sigma_e^4 \left(\sum_{i=0}^{K-1} \sum_{j=-\infty}^{\infty} b_{Kj+i} b_{K(j+h)+i} \right)^2 \\ &\quad + \sum_{r=-\infty}^{\infty} \Big(\gamma_X^2(Kr) + \gamma_X(K(r+h))\gamma_X(K(r-h)) \Big). \end{aligned}$$

This means that the convenient Bartlett formula is not valid in cases of observations of a linear process at lower frequencies. We will come back to this point in Corollary 2.15.

Theorem 2.10 in comparison to Theorem 2.3 implies that in cases in which linear processes (X_t) are not observed at their intrinsic time scale $t = 1, 2, \ldots$ we are confronted with the fact that (asymptotic) variances vary (depending on the frequency K at which we are able to observe the linear process). This means that standard formulas for confidence intervals typically are misleading in such non standard situations.

A similar phenomenon appears for discretely observed stochastic processes in continuous time, cf. Cohen and Lindner (2013) for the situation of discretely observed continuous time moving average processes. So in some sense the situation of discretely observed continuous time moving average processes is comparable to the

situation of discrete time linear processes observed at lower frequencies, as is discussed in the present chapter.

Now we extend Lemma 2.9 and Theorem 2.10 to more general functionals of the periodogram

$$I_n^Y(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} Y_t e^{-i\omega t} \right|^2 , \ \omega \in [0, \pi] , \tag{2.31}$$

of the process (Y_t). Note that the spectral density f_Y of the process (Y_t) (exactly as the autocovariance function, cf. (2.21)) can be represented as a sum of spectral densities of standard linear processes. More precisely we have

$$f_Y = \sum_{i=0}^{K-1} f_Y^{(i)} , \tag{2.32}$$

where $f_Y^{(i)}$ denotes the spectral density of the process $(Y_t^{(i)})$, cf. (2.18).

Theorem 2.13. *Let (X_t) be a linear process and assume that $Ee_t^4 := \eta\, \sigma_e^4 < \infty$ and that $\sum_{j=1}^{\infty} j|b_j| < \infty$. As before the observations are $Y_t = X_{Kt}$, $t = 1, \ldots, n$. Then for every function $W : [0, \pi] \to \mathbb{R}$ of bounded variation $\sqrt{n}(M(I_n^Y, W) - M(f_Y, W))$ converges weakly to a Gaussian distribution with mean zero and variance τ^2 given by*

$$\tau^2 = (\eta - 3) \sum_{i=0}^{K-1} \left(\int_0^{\pi} W(\omega) f_Y^{(i)}(\omega) d\omega \right)^2 + 2\pi \int_0^{\pi} W^2(\omega) \left[\sum_{i=0}^{K-1} f_Y^{(i)}(\omega) \right]^2 d\omega . \tag{2.33}$$

Remark 2.14. *Please note that these results coincide with the ones of Chapter 2.2 if $K = 1$ and thus an appropriate generalization is given. Theorem 2.13 shows that the above results for sample autocovariances transfer to integrated periodograms as well. This means that all above remarks are in force for integrated periodograms.*

Now we turn to ratio statistics, cf. (2.16). The convenient result, that the limiting variance of ratio statistics only depends on second order properties of an underlying linear time series (cf. Corollary 2.6), does not hold any more for linear processes observed at lower frequencies. This is the contents of the following lemma and it means that for general ratio statistics, a class of statistics which contains sample autocorrelations, the well known Bartlett formula is not valid. This is in direct contrast to the result of Corollary 2.6 for linear processes.

Corollary 2.15. *Under the assumptions of Theorem 2.13 and R defined as in (2.12) we obtain that $\sqrt{n}(R(I_n^Y, W) - R(f_Y, W))$ converges to a Gaussian limiting distribution with mean zero and variance*

$$\frac{(\eta-3)}{\left(\int_0^\pi f_Y d\omega\right)^4} \sum_{i=0}^{K-1} \left(\int_0^\pi \left[W \int_0^\pi f_Y d\omega - \int_0^\pi W f_Y d\omega \right] f_Y^{(i)} d\omega \right)^2 \tag{2.34}$$
$$+ \frac{2\pi \int_0^\pi \left[W \int_0^\pi f_Y d\omega - \int_0^\pi W f_Y d\omega \right]^2 \left[\sum_{i=0}^{K-1} f_Y^{(i)} \right]^2 d\omega}{\left(\int_0^\pi f_Y d\omega\right)^4} .$$

As before the second summand in (2.34) coincides with the term obtained for linear processes (cf. (2.16)). Unfortunately, the first summand in (2.34) shows up and typically is not zero. This is the difference to the linear process case. For $K = 1$ the quantities f_Y and $f_Y^{(i)}$ coincide and the first summand vanishes. But in general of course the sum over i and the square can not be interchanged. One exception is given in the Gaussian case. In this case the fourth order cumulant $\eta - 3$ vanishes and therefore we end up with the classical result.

2.4 Estimating and bootstrapping autoregressive processes

Based on the results of Section 2.3 we present a proposal for estimating and bootstrapping parameters of an underlying specific autoregressive model from observations at a lower frequency. We start the discussion with an autoregression of order $p = 2$ and being able to observe only every second data point. Later on, the proposal will be generalized to autoregressive processes of general order p.

2.4.1 Estimating under non-standard observation structure

Let us start with the following example as previously considered in Niebuhr (2011):

Example 2.16. *Consider an autoregressive (AR) process of order $p = 2$, i.e.*

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = e_t \,, \; t \in \mathbb{Z} \,, \tag{2.35}$$

with i.i.d. noise process (e_t), with $Ee_t = 0$, $Ee_t^2 = \sigma_e^2$, for all $t \in \mathbb{Z}$. Assuming the autoregressive polynomial $A(z) = 1 + a_1 z + a_2 z^2$ to have no zeros within the unit circle, (2.35) has an invertible representation as a stationary and causal moving average process of infinite order. Further assume that we only observe $Y_t = X_{2t}$,

i.e. $K = 2$ in the context of Section 2.3. The most frequently used procedure for parameter estimation in AR models is the Yule-Walker method. To apply the Yule-Walker method in order to estimate the parameters a_1 and a_2 of the underlying process we would need sample autocorrelations (of the process (X_t)) at lags $h = 1$ and $h = 2$. From $Y_1, \dots Y_n$ we easily get an estimator for $\gamma_X(0) = \gamma_Y(0)$ and $\gamma_X(2) = \gamma_Y(1)$. But $\gamma_X(1)$ can in principle not be estimated from observations $Y_1, \dots Y_n$. To circumvent the problem we apply a specific linear filter to (2.35). To construct this filter denote by R_1 and R_2 the possibly complex roots of the autoregressive polynomial $A(z)$. One obtains $A(z) = 1 + a_1 z + a_2 z^2 = (1 - z/R_1)(1 - z/R_2)$. We then apply the linear filter $(1 + B/R_1)(1 + B/R_2)$ (where B denotes the backshift operator) to (2.35) and obtain

$$\begin{aligned} &\left(1 + \frac{B}{R_1}\right)\left(1 + \frac{B}{R_2}\right)(X_t + a_1 X_{t-1} + a_2 X_{t-2}) \qquad (2.36)\\ &= \left(1 + \frac{B}{R_1}\right)\left(1 + \frac{B}{R_2}\right)\left(1 - \frac{B}{R_1}\right)\left(1 - \frac{B}{R_2}\right) X_t \\ &= \left(1 + \frac{B}{R_1}\right)\left(1 + \frac{B}{R_2}\right) e_t \,. \end{aligned}$$

Using binomial formula (2.36) leads us to the following ARMA(4,2) model

$$X_t + (2a_2 - a_1^2) X_{t-2} + a_2^2 X_{t-4} \;=\; e_t - a_1 e_{t-1} + a_2 e_{t-2} \,. \qquad (2.37)$$

Switching from X_t to Y_t, (2.37) leads to

$$Y_t + (2a_2 - a_1^2) Y_{t-1} + a_2^2 Y_{t-2} \;=\; \widetilde{\varepsilon}_t \,, \qquad (2.38)$$

where $\widetilde{\varepsilon}_t = e_{2t} - a_1\, e_{2t-1} + a_2\, e_{2t-2}$. Obviously $\widetilde{\varepsilon}_t$ is a 1-dependent noise sequence in the Y_t-sense. Since every p-dependent noise can be represented by a moving average process of order p, we obtain for a properly chosen white noise sequence (ε_t) and a certain coefficient $b \in \mathbb{R}$ from (2.38)

$$Y_t + (2a_2 - a_1^2) Y_{t-1} + a_2^2 Y_{t-2} \;=\; \varepsilon_t - b\, \varepsilon_{t-1} \,. \qquad (2.39)$$

It is noted that (ε_t) is an uncorrelated white noise, but in almost all cases (ε_t) is neither independent nor a martingale difference array. Nevertheless, Yule-Walker estimation is possible from model (2.39) by using observations Y_t only. It should be noticed that the representation (2.39) without odd observations is gained at the cost of losing the independence of the noise, which is a significant price to pay! It is of relevance that the representation (2.37) is not unique in the sense that following

the described way for different autoregressions of order 2 *this may lead to the same autoregressive part of the ARMA(2,1) equation (2.39). More precisely in the case of two complex roots we cannot distinguish between autoregressions of order* 2 *with roots* R_1, R_2 *and with roots* $(-Re(R_1), Im(R_1)), (-Re(R_2), Im(R_2))$. *In the case of two real roots all four combinations of AR(2) processes with roots* $(\pm R_1, \pm R_2)$ *cannot be distinguished.*

The idea of Example 2.16 can be generalized to autoregressive processes of order p. In the setting $K = 2$ the described filtering transforms autoregressions of order p to ARMA$(2p, p)$-structures in X_t-notation. By changing to Y_t-notation this leads to AR(p) representations with a $\lfloor \frac{p}{2} \rfloor$-dependent noise sequence and therefore to ARMA$(p, \lfloor \frac{p}{2} \rfloor)$-structures for (Y_t). Moreover the proposal can be extended to observation lags $K = 2^m$, where $m \in \mathbb{N}_0$.

For further possible extensions to general observation lags K for the process X_t, the main idea would be to find a linear filter, which transforms the autoregressive part in such a way that it can be written in terms of Y_t, only. This means that only observations X_t, with t a multiple of K, appear (compare to (2.37) from above). The right hand side of the filtered AR-equation then always constitutes an m-dependent sequence for some m. Thus we are lead to an ARMA-structure with uncorrelated (but not independent!) white noise. For such models one may apply standard estimation techniques based on moments and sample autocorrelations. Altogether observing linear processes with i.i.d. noise at lower frequencies seems to correspond to observing processes with uncorrelated but dependent noise structure.

Now we come to the main and important disadvantage of this proposal. The described way allows for estimation of the parameters of the transformed (i.e. filtered) processes. Since one mainly is interested in the parameters of the underlying autoregressive process the question is whether or not we can infer back to these quantities. Generally speaking the answer is in the negative as we have already seen for the simplest situation in Example 2.16. To say it more favorably we need to restrict the parameter space for the underlying autoregressive model in such a way that a unique back inference from the final ARMA-model to the underlying autoregressive model gets possible. An example is given by restricting the parameter space in such a way, that the zeroes of the autoregessive polynomial are restricted to one of the complex half planes ($Re(z) > 0$ or $Re(z) < 0$). Then a unique combination of zeroes

would remain, only.

We will see in the next subsection that the asymptotic distribution of sample autocovariances of second order autoregressions for which only $Y_t = X_{2t}$ is observed coincide for both parameter restrictions. Though a back inference to the underlying AR(2)-model is impossible, a consistent bootstrap procedure for the process (Y_t) can be set up.

2.4.2 Bootstrapping AR(2) processes observed at lower frequencies

We restrict to the situation $K = 2$. Based on the discussion in Subsection 2.4.1, we propose a bootstrap procedure for autoregressive models similar to Kreiss (1997) following the lines of Niebuhr (2011). The bootstrap proposal is presented in the situation of Example 2.16. Autoregressive orders higher than 2 will yield similar results but the complexity in notation will be considerably higher. The situation is assumed to be as in Example 2.16. By using the one-sided MA(∞)-representation $X_t = \sum_{j=0}^{\infty} b_j e_{t-j}$, one obtains (by induction, cf. Niebuhr (2011), Corollary 5.15) the coefficients b_j as functions of the AR-parameters

$$b_j = \sum_{i=0}^{j} \binom{j-i}{j-2i} (-a_1)^{j-2i} (-a_2)^i \,, \tag{2.40}$$

where $\binom{\bullet}{a} = 0$ for $a < 0$. Further, using (2.40) the covariance (2.25) can be written as follows

$$\begin{aligned} & c_Y(h,k) \qquad\qquad (2.41)\\ & = \sigma_e^4(\eta-3) \Bigg[\sum_{\nu=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \sum_{i=0}^{2\nu} \sum_{j=0}^{2r} \sum_{t=0}^{2\nu+2h} \sum_{s=0}^{2r+2k} \\ & \quad \Bigg[\binom{2\nu-i}{2\nu-2i} \binom{2\nu+2h-t}{2\nu+2h-2t} \binom{2r-j}{2r-2j} \binom{2r+2k-s}{2r+2k-2s} \\ & \quad + (-a_1)^4 \binom{2\nu+1-i}{2\nu+1-2i} \binom{2\nu+1+2h-t}{2\nu+1+2h-2t} \binom{2r+1-j}{2r+1-2j} \\ & \quad \cdot \binom{2r+1+2k-s}{2r+1+2k-2s} \Bigg] (-a_1)^{4\nu+2h+4r+2k-2i-2t-2j-2s} (-a_2)^{i+t+j+s} \Bigg] \\ & \quad + \sum_{r=-\infty}^{\infty} \left(\gamma_Y(r)\gamma_Y(k-h+r) + \gamma_Y(r+k)\gamma_Y(r-h) \right). \end{aligned}$$

Thus it appears that $|a_1|$ and a_2 together with the second and fourth order moments of the underlying i.i.d. white noise process uniquely determine $c_Y(h,k)$ (note that $\gamma_Y(r)$ only depends on these terms (cf. (2.21))). The autoregressive coefficients can be represented in terms of the roots of the autoregressive polynomial.

$$a_1 = -\left(\frac{1}{R_1}+\frac{1}{R_2}\right) = -\left(\frac{2Re(R_1)}{R_1R_2}\right), \tag{2.42}$$

$$a_2 = \frac{1}{R_1R_2} = \frac{1}{Re^2(R_1)+Im^2(R_1)}. \tag{2.43}$$

Under the assumption of two complex roots we again see that the two AR(2)-models with roots $R_1, R_2 = \overline{R_1}$ or with roots $(-Re(R_1), Im(R_1)), (-Re(R_1), -Im(R_1))$ lead to identical $|a_1|$ and a_2 (cf. (2.42) and (2.43)). It follows that the asymptotic distribution of sample autocovariances for the two not distinguishable AR(2)-models coincide. This opens the door for application of the bootstrap since it asymptotically does not matter which of the two AR(2)-models will be imitated in the bootstrap world. This fact allows us to setup the following valid bootstrap proposal for AR(2) processes having two complex roots of the autoregressive polynomial.

Step 1:
We make use of (2.38) and the definition of $\widetilde{\varepsilon}_t$ below (2.38) to derive from a kind of Yule-Walker equations estimators $\widehat{r}$ and $\widehat{s}$ for the autoregressive coefficients $(2a_2-a_1^2)$ and a_2^2 in (2.38). The explicit formulas for these estimators are

$$\widehat{s} = \frac{\widehat{\gamma}_Y(2) - \frac{\widehat{\gamma}_Y(3)\widehat{\gamma}_Y(1)}{\widehat{\gamma}_Y(2)}}{\frac{\widehat{\gamma}_Y(1)^2}{\widehat{\gamma}_Y(2)} - \widehat{\gamma}_Y(0)}, \qquad \widehat{r} = \frac{\widehat{\gamma}_Y(3)+\widehat{s}\widehat{\gamma}_Y(1)}{-\widehat{\gamma}_Y(2)}. \tag{2.44}$$

Next we compute all four roots of the polynomial $1+\widehat{r}z^2+\widehat{s}z^4$, compare with (2.37). We denote by $\widehat{R}_1$ and $\widehat{R}_2 = \overline{\widehat{R}_1}$ the two conjugate complex roots with positive real parts (the remaining two conjugate complex roots have negative real parts). Having this we define estimators $\widehat{a}_1$ and $\widehat{a}_2$ according to

$$\left(1-z/\widehat{R}_1\right)\left(1-z/\widehat{R}_2\right) = 1+\widehat{a}_1z+\widehat{a}_2z^2. \tag{2.45}$$

These are $\sqrt{n}$-consistent estimators for the AR-parameters $|a_1|$ and a_2. Note that we cannot recover the sign of $|a_1|$, since the true underlying coefficients a_1 and a_2 may belong to the case of complex roots with negative real parts. However, we have seen above that the asymptotic distribution of sample autocovariances does not depend on the sign of a_1, cf. (2.41).

Step 2:
The second and the fourth moment of the white noise e_t consistently can be estimated by

$$\widehat{\sigma}_e^2 := \frac{\widehat{\gamma}_Y(0) + \widehat{r}\widehat{\gamma}_Y(1) + \widehat{s}\widehat{\gamma}_Y(2)}{1 + \widehat{a}_1^2 + \widehat{a}_1^2\widehat{a}_2 - \widehat{a}_2^2} \tag{2.46}$$

and

$$\widehat{E[e_1^4]} := \frac{\frac{1}{n-2}\sum_{t=3}^{n}(Y_t + \widehat{r}Y_{t-1} + \widehat{s}Y_{t-2})^4 - 6\left(\widehat{\sigma_e^2}\right)^2(\widehat{a}_1^2 + \widehat{a}_2^2 + \widehat{a}_1^2\widehat{a}_2^2)}{1 + \widehat{a}_1^4 + \widehat{a}_2^4}\,. \tag{2.47}$$

An estimator of η now reads: $\widehat{\eta} := \widehat{E[e_1^4]} \big/ \left(\widehat{\sigma_e^2}\right)^2$.

Step 3:
In order to generate bootstrap pseudo residuals we define the following discrete distribution with support (x_1, x_2, x_3) and probabilities $(p_1, p_2, 1 - p_1 - p_2)$.

$$x_1 = -x_2 = \sqrt{\widehat{\sigma_e^2}\,\widehat{\eta}} \quad \textit{and} \quad x_3 = 0\,, \tag{2.48}$$

$$p_1 = p_2 = \frac{1}{2\sqrt{\widehat{\eta}}}\,. \tag{2.49}$$

This distribution imitates all relevant moments of the underlying noise sequence up to order four. Assume (e_t^*) to be i.i.d. with this distribution.

Step 4:
In a final step we generate bootstrap pseudo observations $X_1^*, \ldots, X_n^*$ from the following AR(2)-model

$$X_t^* + \widehat{a}_1 X_{t-1}^* + \widehat{a}_2 X_{t-2}^* = e_t^*\,,\ t = 1, 2, \ldots\,, \tag{2.50}$$

which is a linear process. By selecting every second realization of X_t^* we obtain $Y_t^* = X_{2t}^*$, $t = 1, \ldots, n$, as a bootstrap sample of the Y-process.

The following theorem, establishing the consistency of our bootstrap proposal, can be shown.

Theorem 2.17. *Let (X_t) be an autoregressive process of order 2 as in (2.35). The available observations are $Y_t = X_{2t}$ for $t = 1, \ldots, n$. Then for the bootstrap sample*

$Y_1^*, \ldots, Y_n^*$ generated as described above we obtain for the bootstrap sample autocovariances

$$\widehat{\gamma}_{Y^*}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} \left(Y_t^* - \overline{Y^*}\right) \left(Y_{t+|h|}^* - \overline{Y^*}\right) , \tag{2.51}$$

as $n \to \infty$ and in probability

$$n\, Cov^*(\widehat{\gamma}_{Y^*}(h), \widehat{\gamma}_{Y^*}(k) | Y_1, \ldots, Y_n) \to c_Y(h,k) \tag{2.52}$$

where $c_Y(h,k)$ is given by (2.25), and

$$\mathcal{L}\Big(\sqrt{n}(\widehat{\gamma}_{Y^*}(h) - \gamma_{Y^*}(h))_{h=0,\ldots,M} \Big| Y_1, \ldots, Y_n\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_M) , \tag{2.53}$$

where $\gamma_{Y^*}(h) = E^*[Y_t^* Y_{t+h}^* | Y_1, \ldots, Y_n]$ and

$$S_M = (c_Y(h,k))_{h,k=0,\ldots,M} . \tag{2.54}$$

2.4.3 A simulation study

We report here on a simulation study for an autoregressive process of order 2. The process $Y_t = X_{2t}$ is observed and the noise e_t is given by a centered exponential distribution with variance 1. The roots of the autoregressive polynomial are $R_1 = 0.7 + 0.8i$ and $R_2 = 0.7 - 0.8i$. The autoregressive process thus results to

$$\begin{aligned} X_t &= 1.4/1.13\, X_{t-1} - 1/1.13\, X_{t-2} + e_t \\ &\approx 1.239 X_{t-1} + 0.885 X_{t-2} + e_t \end{aligned} \tag{2.55}$$

The sample size is set to $n = 100$, i.e. we have observations $Y_1, \ldots, Y_n$ at hand. The bootstrap procedure of Section 2.4.2 is repeated $B = 1000$ times. And the whole procedure is repeated $M = 1000$ times.

As a quantity of interest of the distribution $\sqrt{n}\,(\widehat{\gamma}_{Y^*}(1) - \widehat{\gamma}_Y(1))$ we choose the 95%-quantile. From the bootstrap we present in Figure 2.1 a boxplot of the obtained bootstrap quantiles. The results are compared with the simulated exact 95% quantile and with the asymptotic 95% quantile from the limiting normal distribution (with estimated variance). Similar results have been obtained for quantiles of other levels and various sample sizes. Figure 2.1 shows that the bootstrap procedure outperforms the normal approximation in this situation.

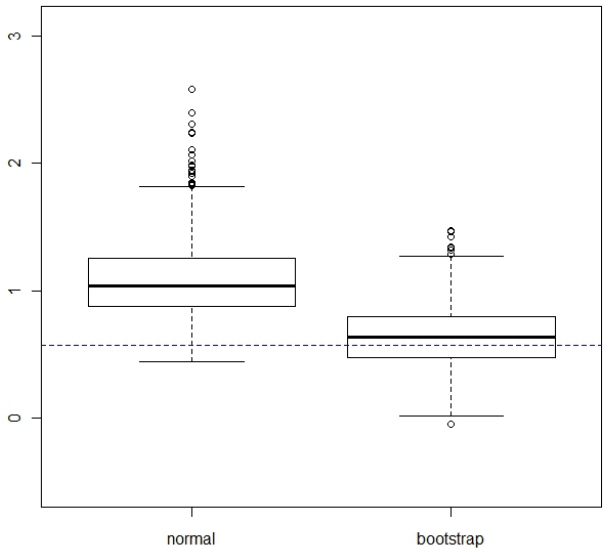

Figure 2.1: 95% quantile of $\sqrt{n}\,(\widehat{\gamma}_Y(1) - \gamma_Y(1))$ of an AR(2) process with sample size $n = 100$. Boxplots of bootstrap approximations and asymptotically normal approximations with estimated variance together with true quantile (dotted line).

2.5 Conclusions

A linear process is assumed to be observable at a lower frequency (mesh size K), only. Depending on this non-standard observation structure it was shown that both the linearity and the independence of the noise is ruined.

Furthermore depending on the observation rhythm K the processes' asymptotic behavior may change for all statistics that can be represented by integrated periodograms. Even the advantage of the transition from integrated periodograms to ratio statistics is lost, since Bartlett's formula generally is not valid. This fact should be taken into account when modeling real valued data with linear processes. Out of this consideration a linear process should not be assumed for a data set if unknown more data points on a finer grid could possibly exist. A rather similar phenomenon appears when discretely observed continuous time processes are considered. Discretely observed continuous time processes will be discussed in the further chapters of the present thesis.

2.6 Proofs

Proof of Lemma 2.9:
The proof is close to the one of Theorem 2.3 part (ii). In a first step we replace $\widehat{\gamma}_Y(h)$ by the following unbiased estimator $\widetilde{\gamma}_Y(h) = n^{-1}\sum_{t=1}^{n}(Y_{t+h}-\mu_Y)(Y_t-\mu_Y)$. Asymptotic equivalence (up to order smaller than n^{-1}) can be shown by straightforward computation. This is in complete analogy to the computation in the proof of Theorem 10.4 in Kreiss and Neuhaus (2006). Thus, the first assertion follows. The computation of the asymptotic covariance of $\widetilde{\gamma}_Y(h)$ now again is straightforward if one uses the representation (2.18). Finally, using the helpful formula $\sum_{s,t=1}^{n} g(s-t) = \sum_{r=-(n-1)}^{n-1}(n-|r|)g(r)$ and applying Kronecker's Lemma the second assertion follows. A detailed proof of this not complicated result is given in Niebuhr (2011), Theorem 4.2. □

Proof of Theorem 2.13:
By using the symmetry of $\overline{\gamma}_n^Y(h) = \frac{1}{n}\sum_{t=1}^{n-|h|} Y_t Y_{t+|h|}$, $|h| < n$, the periodogram simplifies to $I_n^Y(\omega) = \frac{1}{2\pi}\left(\overline{\gamma}_n^Y(0) + 2\sum_{h=1}^{n-1}\overline{\gamma}_n^Y(h)\cos(h\omega)\right)$. By the same reason the spectral density can be written as

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi}\left(\gamma_Y(0) + 2\sum_{h=1}^{\infty}\gamma_Y(h)\cos(h\omega)\right) \\ &= \sum_{i=0}^{K-1}\frac{1}{2\pi}\left(\gamma_Y^{(i)}(0) + 2\sum_{h=1}^{\infty}\gamma_Y^{(i)}(h)\cos(h\omega)\right) \\ &= \sum_{i=0}^{K-1} f_Y^{(i)}(\omega)\,. \end{aligned} \tag{2.56}$$

From $\gamma_Y(h) = \sigma_e^2\sum_{j=0}^{\infty} b_j b_{j+Kh}$ and $\sqrt{n}\,|\sum_{h=n}^{\infty}\gamma_Y(h)\cos(h\omega)| \to 0$, because the coefficients b_j are summable, it suffices to consider

$$\begin{aligned} Z_n := \frac{\sqrt{n}}{2\pi}\Big(&\left(\overline{\gamma}_n^Y(0) - \gamma_Y(0)\right)\int_0^{\pi} W(\omega)d\omega \\ &+2\sum_{h=1}^{n-1}\left(\overline{\gamma}_n^Y(h) - \gamma_Y(h)\right)\int_0^{\pi} W(\omega)\cos(h\omega)d\omega\Big) \end{aligned} \tag{2.57}$$

instead of $\sqrt{n}\Big(M(I_n^Y, W) - M(f_Y, W)\Big)$. Defining, for $M < n$,

$$Z_{n,M} = \sqrt{n}\left(\overline{\gamma}_n^Y(h) - \gamma_Y(h) :\ h = 0,\ldots,M\right)(c_0, c_1, \ldots, c_M)^T\,, \tag{2.58}$$

where $c_0 := \frac{1}{2\pi}\int_0^{\pi} W(\omega)d\omega$, and for $h > 1$, $c_h := \frac{1}{\pi}\int_0^{\pi} W(\omega)\cos(h\omega)d\omega$, we result with a triangular array $Z_{n,M}$. Using Brockwell and Davis (1991), Proposition 6.3.9, the assertion follows if the following three conditions can be verified:

(i) $Z_{n,M} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_M^2)$ as $n \to \infty$, $\forall M \in \mathbb{N}$

(ii) $\tau_M^2 \to \tau^2$ as $M \to \infty$

(iii) $\lim_{M\to\infty} \limsup_{n\to\infty} P\left(|Z_n - Z_{n,M}| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0\,.$

The first condition (i) directly follows by applying Theorem 2.10. Furthermore, using $S_M = \left(c_Y(h,k)\right)_{h,k=0,\ldots,M}$ and denoting $\tau_M^2 := (c_0, \ldots, c_M) S_M (c_0, \ldots, c_M)^T$ one obtains

$$\begin{aligned} \tau_M^2 &= (\eta - 3) \sum_{i=0}^{K-1} \left(\sum_{h=0}^{M} c_h \gamma_Y^{(i)}(h) \right)^2 \\ &\quad + \sum_{r=-\infty}^{\infty} \sum_{h,k=0}^{M} c_h\, c_k \left(\gamma_Y(r)\gamma_Y(r+k-h) + \gamma_Y(r+k)\gamma_Y(r-h) \right). \end{aligned} \tag{2.59}$$

Absolute summability of c_h and $\gamma_Y(h)$ yield

$$\begin{aligned} &\int_0^\pi W(\omega) f_Y^{(i)}(\omega) d\omega \\ &= \int_0^\pi W(\omega) \frac{1}{2\pi} \left(\gamma_Y^{(i)}(0) + 2 \sum_{h=1}^{\infty} \gamma_Y^{(i)}(h) \cos(h\omega) \right) d\omega \\ &= \sum_{h=0}^{\infty} c_h \gamma_Y^{(i)}(h) = \sum_{h=0}^{\infty} c_h \gamma_Y^{(i)}(h) \end{aligned} \tag{2.60}$$

because of (2.56). From this it follows that the first summand of τ_M^2 converges to the corresponding first summand of τ^2 (cf. (2.33) for the definition of τ^2). In order to obtain the same assertion for the second summand we make use of $W(\omega) = 2\sum_{h=0}^{\infty} c_h \cos(h\omega)$ and the following addition formulae for trigonometric functions

$$\begin{aligned} &4\, \cos(a)\cos(b)\cos(c) \\ &= \cos(a+b-c) + \cos(b+c-a) + \cos(c+a-b) + \cos(a+b+c)\,. \end{aligned} \tag{2.61}$$

We obtain

$$\begin{aligned} &2\pi \int_0^\pi W^2(\omega) \left(\sum_{i=0}^{K-1} f_Y^{(i)}(\omega) \right)^2 d\omega = 2\pi \int_0^\pi W^2(\omega) f_Y^2(\omega) d\omega \\ &= 4 \sum_{h,k=0}^{\infty} c_h\, c_k \int_0^\pi \cos(h\omega)\cos(k\omega) \left(\sum_{r=-\infty}^{\infty} \gamma_Y(r) \cos(r\omega) f_Y(\omega) \right) d\omega \\ &= \frac{1}{2} \sum_{h,k=0}^{\infty} c_h\, c_k \sum_{r=-\infty}^{\infty} \left(\gamma_Y(r)\gamma_Y(r+k-h) + \gamma_Y(r)\gamma_Y(k+h-r) \right. \\ &\qquad \left. + \gamma_Y(r)\gamma_Y(r+h-k) + \gamma_Y(r)\gamma_Y(r+k+h) \right). \end{aligned} \tag{2.62}$$

Now (ii) immediately follows. To proof the third condition (iii) by Markov inequality it suffices to handle $E|Z_n - Z_{n,M}|$. Observe

$$E\,|Z_n - Z_{n,M}| \leq \sum_{h=M+1}^{n-1} \sqrt{n}\,|c_h|\,E\left|\overline{\gamma}_n^Y(h) - \gamma_Y(h)\right| \tag{2.63}$$

and obtain by direct computation that

$$\left(\sqrt{n}E|\overline{\gamma}_n^Y(h) - \gamma_Y(h)|\right)^2 \leq nE\left(\overline{\gamma}_n^Y(h) - \gamma_Y(h)\right)^2 \tag{2.64}$$

is uniformly bounded in h. Absolute summability of c_h leads to assertion (iii). □

Proof of Theorem 2.17:
The proof follows the lines of Niebuhr (2011). Due to the stepwise bootstrap proposal (cf. (2.50)) the process Y_t^* has a one-sided moving average representation, namely

$$Y_t^* = \sum_{\nu=0}^{\infty} \widehat{b}_\nu e_{t-\nu}^*,\ t \in \mathbb{Z}, \tag{2.65}$$

with $\sum_{\nu=0}^{\infty} \widehat{b}_\nu z^\nu = (1 + \widehat{a}_1 z + \widehat{a}_2 z^2)^{-1}$, $|z| \leq 1$. Convergence of the power series for n large enough is guaranteed because of consistency of $|\widehat{a}_1|$ and $\widehat{a}_2$ for $|a_1|$ and a_2. As a direct consequence one obtains

$$\gamma_{Y^*}(h) = E^*[Y_t^* Y_{t+h}^* | Y_1, \ldots, Y_n] = \widehat{\sigma}_e^2 \sum_{\nu=0}^{\infty} \widehat{b}_\nu \widehat{b}_{\nu+h}. \tag{2.66}$$

In a first step we show that (2.52) holds for $\widetilde{\gamma}_{Y^*}(h) = n^{-1}\sum_{t=1}^n Y_{t+h}^* Y_t^*$ instead of $\widehat{\gamma}_{Y^*}(h)$. We obtain by straightforward computation, and using (2.65) in the third equality,

$$\begin{aligned}
&n\,\mathrm{Cov}^*(\widetilde{\gamma}_{Y^*}(h), \widetilde{\gamma}_{Y^*}(k) | Y_1, \ldots, Y_n) \qquad (2.67)\\
&= n\,[E^*[\widetilde{\gamma}_{Y^*}(h)\widetilde{\gamma}_{Y^*}(k)|Y_1, \ldots, Y_n] - E^*[\widetilde{\gamma}_{Y^*}(h)|Y_1, \ldots, Y_n]E^*[\widetilde{\gamma}_{Y^*}(k)|Y_1, \ldots, Y_n]]\\
&= n\,[E^*[\widetilde{\gamma}_{Y^*}(h)\widetilde{\gamma}_{Y^*}(k)|Y_1, \ldots, Y_n] - \gamma_{Y^*}(h)\gamma_{Y^*}(k)]\\
&= \widehat{\sigma}_e^4(\widehat{\eta} - 3)\frac{1}{n}\sum_{i=0}^{K-1}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\nu=-\infty}^{\infty} \widehat{b}_\nu \widehat{b}_{\nu+ih}\widehat{b}_{\nu+i(s-t)}\widehat{b}_{\nu+i(s-t+k)}\\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}[\gamma_{Y^*}(s-t)\gamma_{Y^*}(k+s-t-h) + \gamma_{Y^*}(s-t+k)\gamma_{Y^*}(s-t-h)]\\
&= \sum_{r=-(n-1)}^{n-1}\left(1 - \frac{|r|}{n}\right)T_r^*,
\end{aligned}$$

where

$$\begin{aligned} T_r^* &= \widehat{\sigma}_e^4(\widehat{\eta}-3)\sum_{i=0}^{K-1}\sum_{\nu=-\infty}^{\infty}\widehat{b}_\nu\widehat{b}_{\nu+ih}\widehat{b}_{\nu+ir}\widehat{b}_{\nu+i(r+k)} \\ &\quad +\gamma_{Y^*}(r)\gamma_{Y^*}(k+r-h)+\gamma_{Y^*}(r+k)\gamma_{Y^*}(r-h). \end{aligned} \tag{2.68}$$

Using Cauchy's inequality for holomorphic functions (compare Kreiss and Franke (1992), Lemma 2.2), we immediately obtain

$$\sup_{\nu\in\mathbb{N}}(1+\delta)^\nu\left|\widehat{b}_\nu-b_\nu\right|=\mathcal{O}_P(n^{-1/2}) \tag{2.69}$$

for some $\delta>0$. Equation (2.69) together with the consistencies of $\widehat{\sigma}_e^4$ and $\widehat{\eta}$ for σ_e^4 and η (both are immediate consequences of the weak law of large numbers), respectively, as well as the absolute summability of the coefficients b_ν now lead in a direct but tedious computation to the convergence

$$\sum_{r=-(n-1)}^{n-1}|T_r^*-T_r|=o_P(1)\,, \tag{2.70}$$

where T_r is defined exactly as T_r^*, cf. (2.68), with $\widehat{\sigma}_e^2$, $\widehat{\eta}$, $\widehat{b}_\nu$ and γ_{Y^*} replaced by σ_e^2, η, b_ν and γ_Y, respectively. This means by (2.67) that

$$n\,\mathrm{Cov}^*(\widetilde{\gamma}_{Y^*}(h),\widetilde{\gamma}_{Y^*}(k)|Y_1,\ldots,Y_n)\to c_Y(h,k). \tag{2.71}$$

Now we show the asymptotic equivalence of $\widetilde{\gamma}_{Y^*}$ and $\widehat{\gamma}_{Y^*}$. It holds

$$\begin{aligned} \widehat{\gamma}_{Y^*}(h) &= \frac{1}{n}\sum_{t=1}^{n-h}\left(Y_{t+h}^*-\overline{Y^*}\right)\left(Y_t^*-\overline{Y^*}\right) \\ &= \frac{1}{n}\sum_{t=1}^{n}Y_{t+h}^*Y_t^*-\frac{1}{n}\sum_{t=n-h+1}^{n}Y_{t+h}^*Y_t^* \\ &\quad -\overline{Y^*}\frac{1}{n}\sum_{t=1}^{n-h}Y_t^*-\overline{Y^*}\frac{1}{n}\sum_{t=1}^{n-h}Y_{t+h}^*+\frac{n-h}{n}\left(\overline{Y^*}\right)^2 \\ &= \widetilde{\gamma}_{Y^*}(h)+R_{n,1}^*+\ldots+R_{n,4}^* \end{aligned} \tag{2.72}$$

In order to obtain assertion (2.52) from (2.71) it suffices to show that for all remainder terms $R_{n,j}^*$, $j=1,\ldots,4$, in (2.72): $Var^*(R_{n,j}^*)=\mathcal{O}_P(n^{-2})$. Exemplary consider $Var^*(R_{n,4}^*)\le\frac{1}{n^4}E^*\left(\sum_{s=1}^n Y_s^*\right)^4$ and use (2.65) and (2.69) to bound this expectation through $\mathcal{O}_\mathcal{P}(n^{-2})$. All other remainder terms can be treated in the same way. Thus, assertion (2.52) follows. This is the first part of Theorem 2.17.

For a proof of the second part of Theorem 2.17 we make use of Brockwell and Davis (1991), Proposition 6.3.9. We define $Y^*_{t,m} = \sum_{j=0}^{m} \widehat{b}_j e^*_{t-j}$. This sequence is an m-dependent triangular array. The CLT for m-dependent triangular arrays gives

$$\sqrt{n}\,(\widehat{\gamma}_{Y^*,m}(0) - \gamma_{Y^*,m}(0), \ldots, \widehat{\gamma}_{Y^*,m}(M) - \gamma_{Y^*,m}(M)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, S_{M,m}), \qquad (2.73)$$

where $\widehat{\gamma}_{Y^*,m}$ and $\gamma_{Y^*,m}$ are defined as $\widehat{\gamma}_{Y^*}$ and γ_{Y^*} with Y^*_t replaced by $Y^*_{t,m}$. $S_{M,m}$ is defined by (2.54) and (2.25) with ∞ replaced by m. Since $S_{M,m} \to S_M$ as $m \to \infty$ and for every $\varepsilon > 0$ and every $h \in \mathbb{N}_0$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left\{|\sqrt{n}(\widehat{\gamma}_{Y^*}(h) - \gamma_{Y^*}(h)) - \sqrt{n}(\widehat{\gamma}_{Y^*,m}(h) - \gamma_{Y^*,m}(h))| > \varepsilon\right\} = 0 \qquad (2.74)$$

the above mentioned result (i.e. Brockwell and Davis (1991), Theorem 6.3.9) yields the second part of Theorem 2.17. □

3 The continuous-time autoregressive moving average model

Classical time series analysis is based on the assumption that the data-generating process is of discrete time. Nevertheless, continuous-time models have also been of interest for many years (see e.g. Fowler (1936), Doob (1944), etc.). Discrete time series are often interpreted as realizations of continuous-time processes. It is then natural, even though the observations are made at discrete times, to model the underlying data-generating process as a continuous time series.

In the last ten years there has been increasing interest in continuous time models which is partly due to the very succesful application of stochastic differential equation models to problems in finance. The most popular candidate is the derivation of the Black-Scholes option-pricing formula and its generalizations (Hull and White (1987)). Further, Jones (1981) exploited continuous time models for dealing with missing values in time series. Then the question for the treatment of irregularly spaced data ensues quite naturally (Jones (1985)).

Among all continuous time models, we will focus on continuous-time autoregressive moving average (CARMA) processes driven by (second-order) Lévy processes. Just as ARMA processes play a central role in time series analysis with discrete time parameter, CARMA processes play an analogous role for processes with continuous time parameter. In this chapter we will define the CARMA process and its usual requirements. The question for bootstrap possibilities for samples stemming from this general class of processes is the main motivation for this thesis. In this chapter we outline some basic definitions and results which will be required later on.

Definition 3.1. *A stochastic process $L = \{L(t)\}_{t\in\mathbb{R}}$ is said to be a Lévy process if the following conditions hold:*

(i) *L has independent and stationary increments.*

(ii) *L is continuous in probability.*

(iii) *L has sample-paths which are right-continuous with left limits (càdlàg).*

(iv) *$L(0) = L(0-) = 0$.*

If L further has the property $EL(1)^2 < \infty$, then L is called a second-order Lévy process.

Example 3.2. *The most famous examples of Lévy processes are the Poisson process, N, for which $N(t)$, $t \geq 0$, has the Poisson distribution with mean λt for some fixed $\lambda > 0$, and the Brownian motion, B, for which $B(t)$, $t \geq 0$, is normally distributed with mean μt and variance $\sigma^2 t$ for some fixed $\mu \in \mathbb{R}$ and $\sigma > 0$.*

Definition 3.3. *We define a Lévy-driven CARMA(p, q) process $\{Y(t)\}$ with $0 \leq q < p$ and coefficients $a_1, \ldots, a_p, b_0, \ldots, b_q$ to be a stationary solution of the p-th order differential equation, called the state-space representation,*

$$a(D)Y(t) = b(D)DL(t), \quad t \geq 0, \tag{3.1}$$

where D denotes differentiation with respect to t, $\{L_t\}$ is a (second-order) Lévy process,

$$a(z) = z^p + a_1 z^{p-1} + \ldots + a_p, \tag{3.2}$$

$$b(z) = b_0 + b_1 z + \ldots + b_p z^p, \tag{3.3}$$

and the coefficients b_j satisfy $b_q \neq 0$ and $b_j = 0$ for $q < j \leq p$. Since the derivatives $DL(t)$ do not exist in the usual sense, we interpret (3.1) as being equivalent to the so-called observation and state equation,

$$Y(t) = \underline{b}^T \underline{X}(t) \tag{3.4}$$

and

$$d\underline{X}(t) - A\underline{X}(t)dt = \underline{e}dL(t), \tag{3.5}$$

where

$$\underline{X}(t) = \begin{bmatrix} X_0(t) \\ X_1(t) \\ \vdots \\ X_{p-2}(t) \\ X_{p-1}(t) \end{bmatrix}, \quad \underline{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{bmatrix} \tag{3.6}$$

and $\underline{X}(0)$ *is assumed to be uncorrelated with* $\{L(t)\}$.

Remark 3.4. *The state equation (3.5) is equivalent to*

$$\underline{X}(t) = e^{A(t-s)}\underline{X}(s) + \int_{(s,t]} e^{A(t-u)}\underline{e}dL(u), \quad \forall s \leq t \in \mathbb{R}. \tag{3.7}$$

This equation will be of essential use in Chapter 4.

Remark 3.5. *It is easy to check that the eigenvalues of the matrix A are the same as the zeroes of the autoregressive polynomial* $a(z)$.

Remark 3.6. *In the case of* $p = 1$ *and* $q = 0$ *the matrix A is to be read as a single coefficient* $A = (-a_1)$. *Indeed, this is the Ornstein-Uhlenbeck case.*

Remark 3.7. *If the driving Lévy process is a second-order Lévy process then the second-order characteristic transfers to Y and it holds* $EY(t)^2 < \infty$. *The process Y then is not only strictly stationary but also weakly stationary (cf. Masuda (2004)).*

Assumption 1. *Assume that the eigenvalues* $\lambda_1, \ldots, \lambda_p$ *of the matrix A satisfy the condition that their real parts are strictly negative, i.e.*

$$Re\,(\lambda_i) < 0 \quad \forall i = 1, \ldots, p. \tag{3.8}$$

Proposition 3.8 (Yang (2007), Proposition 1.3.1). *If* $\{\underline{X}(0)\}$ *is independent of the driving Lévy process* $\{L(t), t \geq 0\}$ *and* $EL(1)^2 < \infty$, *then* $\{\underline{X}(t)\}$ *is strictly stationary if and only if Assumption 1 holds and* $\{\underline{X}(t)\}$ *has the distribution of* $\int_0^\infty e^{-Au}\underline{e}dL(u)$.

Necessary and sufficient conditions for the existence and uniqueness for solutions of stationary Lévy-driven CARMA processes were stated by Brockwell and Lindner (2009). In their Theorem 4.1 it is shown that without loss of generality it can be assumed that the polynomials $a(z)$ and $b(z)$ have no common zeroes. If that

is the case and if some process is a solution of the observation and state equation, it will also satisfy the equations after cancelling the common factors of $a(z)$ and $b(z)$.

For later discussions we restate some further general results on CARMA processes from Brockwell et al. (2010).

Proposition 3.9 (Brockwell et al. (2010), Proposition 2.). *Let Assumption 1 hold, then the CARMA(p, q) process can expressed as a sum of dependent and possibly complex-valued CAR(1) processes. Specifically*

$$Y(t) = \sum_{r=1}^{p} Y^{(r)}(t), \tag{3.9}$$

where

$$Y^{(r)}(t) \;=\; \int_{-\infty}^{t} \alpha_r e^{\lambda_r (t-u)} dL(u), \tag{3.10}$$

and

$$\alpha_r \;=\; \frac{b(\lambda_r)}{a'(\lambda_r)}, \;\; r = 1, \ldots, p, \tag{3.11}$$

and a' denotes the derivative of the function a.

Proposition 3.10 (Brockwell et al. (2010), Proposition 3.). *Under Assumption 1, the sampled process $\{Y_t := Y(th)\}$, where $\{Y(t)\}$ is defined as above, is given by $Y_t = \sum_{r=1}^{p} Y_t^{(r)}$, where the discrete-time processes $\left\{Y_t^{(r)}\right\}$, $r = 1, \ldots, p$, are obtained by sampling the component CAR(1) processes $\left\{Y^{(r)}(t)\right\}$ at spacing h. Thus*

$$Y_t^{(r)} = e^{\lambda_r h} Y_{t-1}^{(r)} + Z_t^{(r)}, \;\; t = 0, \pm 1, \ldots, \tag{3.12}$$

with i.i.d. noise

$$Z_t^{(r)} = \alpha_r \int_{(t-1)h}^{th} e^{\lambda_r (th-u)} dL(u). \tag{3.13}$$

The sampled process $\{Y_t\}$ satisfies the equations,

$$\Phi(B) Y_t = V_t^1 + V_{t-1}^2 + \ldots + V_{t-p+1}^p, \tag{3.14}$$

where

$$\Phi(z) := \prod_{r=1}^{p} (1 - e^{\lambda_r h} z) =: 1 - \phi_1 z - \ldots - \phi_p z^p \tag{3.15}$$

and $B^j Y_t := Y_{t-j}$*. For each fixed* $r \in \{1, \ldots, p\}$*,* $\{V_t^r\}$ *is the i.i.d. sequence defined by*

$$V_t^r = \int_{(t-1)h}^{th} \sum_{k=1}^{p} \alpha_k \left(e^{(r-1)h\lambda_k} - \sum_{j=1}^{r-1} \phi_j e^{(r-1-j)h\lambda_k} \right) e^{(th-u)\lambda_k} dL(u). \tag{3.16}$$

The right-hand side of (3.14) can also be expressed as an invertible moving average

$$\Theta(B)W_t := W_t + \theta_1 W_{t-1} + \ldots + \theta_{p-1} W_{n-p+1}, \tag{3.17}$$

where $\{W_t\}$ *is a white (not necessarily i.i.d.) noise sequence and* $\theta_1, \ldots, \theta_{p-1}$ *are constants depending on the parameters of the CARMA process. Thus* $\{Y_t\}$ *is the weak ARMA*$(p, p-1)$ *process satisfying*

$$\Phi(B)Y_t = \Theta(B)W_t, \tag{3.18}$$

where

$$W_t = \Theta(B)^{-1} \sum_{r=1}^{p} V_{t-r+1}^r. \tag{3.19}$$

The importance of these results for our investigations lies especially in equation (3.18). Any (equidistantly) sampled CARMA process has an ARMA representation. If Y is a Gaussian CARMA process, then the sampled process is found to be a discrete-time autoregressive moving average (ARMA(r, s)) process, with $0 \leq s < r \leq p$ driven by independent and identically distributed Gaussian white noise. This characteristic is charming for the bootstrap to settle on. However, samples of general Lévy-driven CARMA processes are also ARMA processes but with a driving noise sequence being not necessarily i.i.d.. This is a crucial difference for the ability to apply bootstrap approaches. Such weak ARMA processes, but also AR processes with some dependent noise, are not straightforwardly to bootstrap. Basicly, this is what motivates main parts of the remainder of this thesis, namely the bootstrap for continuous-time autoregressions as well as the autoregressive-aided block bootstrap.

4 | Bootstrapping continuous-time autoregressive processes

Based on : Brockwell, P. J., Kreiß, J.-P. and Niebuhr, T. (2014):
Bootstrapping continuous-time autoregressive processes.
The Annals of the Institute of Statistical Mathematics, Vol. 66, pp. 75–92.

Abstract. We develop a bootstrap procedure for Lévy-driven continuous-time autoregressive (CAR) processes observed at discrete regularly-spaced times. It is well-known that a regularly sampled stationary Ornstein-Uhlenbeck process (i.e. a CAR process of order 1) has a discrete-time autoregressive representation with i.i.d. noise. Based on this representation a simple bootstrap procedure can be found. Since regularly sampled CAR processes of higher order satisfy ARMA equations with uncorrelated (but in general dependent) noise, a more general bootstrap procedure is needed for such processes. We consider statistics depending on observations of the CAR process at the uniformly-spaced times, together with auxiliary observations on a finer grid, which give approximations to the derivatives of the continuous time process. This enables us to approximate the state-vector of the CAR process which is a vector-valued CAR(1) process, and whose sampled version, on the uniformly-spaced grid, is a multivariate AR(1) process with i.i.d. noise. This leads to a valid residual-based bootstrap which allows replication of CAR(p) processes on the underlying discrete time grid. We show that this approach is consistent for empirical autocovariances and autocorrelations.

4.1 Introduction

The modeling of continuous time processes has a long history and has been carried out widely in financial econometrics. Early papers of Doob (1944) and Phillips (1959) deal with representations and properties of Gaussian continuous-time ARMA processes. State-space representations of these processes were exploited by Jones (1981) for dealing with missing values in time series, and by Brockwell (2001) in the study of Lévy-driven continuous time ARMA (CARMA) processes. These allow the modeling of series with a wide variety of marginal distributions including heavy-tailed and asymmetric distributions. Long-memory versions have been developed by Brockwell and Marquardt (2005). One of the important applications of Lévy-driven CARMA processes is in financial econometrics where they have been used as models for spot volatility in stochastic volatility models (Barndorff-Nielsen and Shephard (2001), Brockwell and Lindner (2012)). Over the years, the topic of embedding a discrete-time ARMA process in a continuous-time ARMA process has also been studied by a number of authors including Chan and Tong (1987), He and Wang (1989), Huzii (2001) and Brockwell (1994), (2009), whose results will be important for our work later in this chapter. High-frequency sampling of CARMA processes has also been studied by Brockwell et al. (2012) in connection with the extremely high-frequency measurements of turbulent wind-speed which are available now. The bootstrap possibilities for this huge class of processes have not previously been investigated.

This chapter of the present thesis is concerned with bootstrapping statistics of general Lévy-driven CAR processes on general but fixed time grids with spacing $\Delta > 0$. We define the CAR(p) process in Section 4.2 and give an overview of its representations. This will clarify the agreements and the differences from the general CARMA model as defined in Chapter 3. Afterwards, we will briefly review the results of Cohen and Lindner (2013) who handle equidistant samples of continuous-time moving average processes and give another representation for the limiting variance in their central limit result. This representation will be helpful for proving an analogous bootstrap result later in this chapter. Section 4.3 discusses the Ornstein-Uhlenbeck case and its special characteristics. The bootstrapping of Ornstein-Uhlenbeck processes reduces to a very simple and well-known situation. Unfortunately, this is not the case for higher-order CAR processes. A bootstrap procedure for the general case is then proposed and investigated, concluding with a simulation study in Section 4.4.

4.2 The continuous-time autoregressive model

We define a second-order Lévy-driven CAR(p) process $\{Y(t)\}$ with $p > 0$ and parameters $a_1, \ldots, a_p$ to be a stationary solution of the formal differential equation

$$a(D)Y(t) = DL(t), \quad t \geq 0, \tag{4.1}$$

where D denotes differentiation with respect to t, L is a second order Lévy process, and the polynomial $a(z)$ is defined by

$$a(z) \;=\; z^p + a_1 z^{p-1} + \ldots + a_p \,. \tag{4.2}$$

Since the derivative of the Lévy process $L(t)$ does not exist in the usual sense, we follow the standard approach via the state-space representation of (4.1) (cf. Brockwell (2009) for an overview),

$$Y(t) \;=\; \underline{b}^T \underline{X}(t) = X_0(t), \tag{4.3}$$

$$d\underline{X}(t) - A\underline{X}(t)dt \;=\; \underline{e}\,dL(t), \tag{4.4}$$

where

$$\underline{X}(t) = \begin{bmatrix} X_0(t) \\ X_1(t) \\ \vdots \\ X_{p-2}(t) \\ X_{p-1}(t) \end{bmatrix}, \; \underline{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \; \underline{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{4.5}$$

and

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \ldots & -a_1 \end{bmatrix}. \tag{4.6}$$

Note that (4.4) is a system of p stochastic differential equations. Except for the last one, all equations are of the same type and give

$$X_j = X_0^{(j)}, \quad j = 0, \ldots, p-1. \tag{4.7}$$

Thus, the components of $\underline{X}$ are the derivatives of the CAR process $Y = X_0$. Every solution of (4.4) satisfies the equation,

$$\underline{X}(t) \;=\; e^{A(t-s)}\underline{X}(s) + \int_s^t e^{A(t-u)}\underline{e}\,dL(u) \;\; \forall s < t. \tag{4.8}$$

Remark 4.1. *Definition 3.3 is in line with the above description of a CAR process when the moving average coefficients are treated as zeroes.*

For the existence of a weakly stationary and causal solution $(Y(t))$ of the equations (4.3) and (4.4) under the assumption $E\,L(1)^2 < \infty$ it is necessary and sufficient that the zeroes $\lambda_1, \lambda_2, \ldots, \lambda_p$ of (4.2) (which coincide with the eigenvalues of the matrix A) all have strictly negative real parts (see Brockwell (2001) or Brockwell (2009), Proposition 1). Under these assumptions, the solution is also strictly stationary (cf. Brockwell (2009), Proposition 2). The stationary solution of (4.8) is given by

$$\underline{X}(t) \;=\; \int_{-\infty}^{t} e^{A(t-u)}\underline{e}\, dL(u)\;, \tag{4.9}$$

while the unique weakly and strictly stationary solution of (4.3) and (4.4) reads

$$Y(t) \;=\; \underline{b}^T\underline{X}(t) = \int_{-\infty}^{t} \underline{b}^T e^{A(t-u)}\underline{e} dL(u) = \int_{-\infty}^{\infty} f(t-u) dL(u)\,. \tag{4.10}$$

In (4.10), the function $f(t) = \underline{b}^T e^{At}\underline{e}\mathbf{1}_{[0,\infty)}(t)$ is referred to as the kernel of the CAR process $Y(t)$, see e.g. Brockwell et al. (2010) or Cohen and Lindner (2013). This is the reason for making the following assumptions.

Assumption 2. *The zeroes $\lambda_1, \lambda_2, \ldots, \lambda_p$ of the autoregressive polynomial (4.2) (which are also the eigenvalues of the matrix A) are all assumed to have strictly negative real parts.*

Assumption 3. *The driving Lévy process is assumed to have zero mean, variance $\sigma^2 := E\,L(1)^2 < \infty$ and $\eta := \sigma^{-4}\,E\,L(1)^4 < \infty$.*

Remark 4.2. *Although $Y = X_0$ is a univariate process, the state representation (4.3) and (4.4) characterizes it as the first component of a multivariate state-vector $\underline{X}$, i.e. for $\underline{b}^T = (1, 0, \ldots, 0)$ we have $Y(t) = \underline{b}^T\underline{X}(t)$. This leads us to*

$$\varrho_Y(q) = \frac{\gamma_Y(q)}{\gamma_Y(0)} = \frac{\underline{b}^T\underline{\mathbf{\Gamma}}(q)\underline{b}}{\underline{b}^T\underline{\mathbf{\Gamma}}(0)\underline{b}}\,, \tag{4.11}$$

where $\underline{\mathbf{\Gamma}}(q)$ denotes the autocovariance matrix of $\underline{X}$ at lag q.

For the estimation and bootstrap procedure we assume that high-frequency observations are available for the estimation of certain derivatives while our interest lies in the behaviour of the process on a fixed Δ-grid. More detailed comments on the

observation structure are given later. For technical reasons in the proof of the ensuing bootstrap procedure we present another representation of the CAR process $Y(t)$ itself. Using (4.8) with $t\Delta$ and $(t+1)\Delta$ as bounds of the integral, we obtain

$$\underline{X}((t+1)\Delta) \;=\; e^{A\Delta}\underline{X}(t\Delta) + \int_{t\Delta}^{(t+1)\Delta} e^{A((t+1)\Delta-u)}\underline{e}dL(u)\,, \tag{4.12}$$

that is a vector autoregressive representation (VAR) of order one. Abbreviating the i.i.d. noise sequence by

$$\underline{Z}((t+1-j)\Delta) := \int_{(t-j)\Delta}^{(t+1-j)\Delta} e^{A((t+1)\Delta-u)}\underline{e}dL(u)\,,\quad t\in\mathbb{Z}\,, \tag{4.13}$$

and inverting the VAR(1)-representation (4.12) leads to the following moving average representation of the process $(\underline{X}(t\Delta) : t \in \mathbb{Z})$:

$$\underline{X}((t+1)\Delta) \;=\; \sum_{j=0}^{\infty}\left(e^{A\Delta}\right)^j \underline{Z}((t+1-j)\Delta)\,. \tag{4.14}$$

Correspondingly, the sampled CAR(p)-process $(Y(t\Delta))$ itself can be written as

$$\begin{aligned} Y((t+1)\Delta) &= \sum_{j=0}^{\infty}\underline{b}^T e^{A\Delta j}\underline{Z}((t+1-j)\Delta) \\ &= \sum_{j=0}^{\infty}\underline{c}_j^T\underline{Z}((t+1-j)\Delta) \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{p-1} c_{j,i}Z_i((t+1-j)\Delta)\,, \end{aligned} \tag{4.15}$$

where $\underline{c}_j^T := \underline{b}^T\left(e^{A\Delta}\right)^j, j\in\mathbb{N}_0$, is a sequence of p-variate coefficients.

Remark 4.3. *It is worth mentioning that the moving average representation (4.15) varies with Δ, since the coefficients depend on the fixed grid size Δ.*

Under the assumption of finite fourth moments and an appropriate Lévy process, Cohen and Lindner (2013) investigated continuous-time moving average processes of infinite order if observations are taken on a fixed Δ-grid. Their Theorem 3.3 gives the asymptotic normal distribution for empirical autocovariances and autocorrelations based on observations taken on a fixed Δ-grid. It is worth mentioning that the asymptotic variance-covariance matrix substantially differs from the matrix obtained for linear processes in discrete time, i.e. discrete time moving average processes of possibly infinite order, and, most important, with i.i.d. innovations (cf. Brockwell and Davis (1991), Proposition 7.3.4 and Theorem 7.2.1). If we specialize Theorem 3.3 of Cohen and Lindner (2013) to CAR(p)-processes we obtain:

Proposition 4.4. *Under Assumptions 2 and 3 we obtain for the empirical autocovariances*

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=0}^{n-h-1}\left(Y_{t+h}-\overline{Y}\right)\left(Y_t-\overline{Y}\right), \quad h=0,\ldots,n-1, \tag{4.16}$$

where

$$\overline{Y} = \frac{1}{n}\sum_{t=0}^{n-1} Y_t\,, \tag{4.17}$$

of observations $(Y_t := Y(t\Delta) : t = 0,\ldots,n-1)$ *stemming from a CAR(p) process sampled on a fixed* Δ*-grid, as* $n\to\infty$*, that*

$$\sqrt{n}\Big(\widehat{\gamma}_Y(0)-\gamma_Y(0),\ldots,\widehat{\gamma}_Y(q)-\gamma_Y(q)\Big)^T \xrightarrow{\mathcal{D}} \mathcal{N}(0,V), \tag{4.18}$$

where the variance-covariance matrix $V = (v_{q_1,q_2,\Delta})_{q_1,q_2=0,\ldots,q} \in \mathbb{R}^{q+1,q+1}$ *has components*

$$\begin{aligned} &v_{q_1,q_2,\Delta} \\ &= \sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1}\sum_{j=0}^{\infty} c_{j,i_1}c_{j+q_1,i_2}\sum_{r=0}^{\infty} c_{j+r,i_3}c_{j+r+q_2,i_4} \\ &\quad\cdot\Big(\kappa^4_{i_1,i_2,i_3,i_4} - E[Z_{i_1}Z_{i_2}]E[Z_{i_3}Z_{i_4}] - E[Z_{i_1}Z_{i_3}]E[Z_{i_2}Z_{i_4}] - E[Z_{i_1}Z_{i_4}]E[Z_{i_2}Z_{i_3}]\Big) \\ &\quad+\sum_{r=-\infty}^{\infty}\Big\{\gamma_Y(r\Delta)\gamma_Y\left((q_2+r-q_1)\Delta\right)+\gamma_Y\left((r+q_2)\Delta\right)\gamma_Y((r-q_1)\Delta)\Big\}, \end{aligned} \tag{4.19}$$

and $\kappa^4_{i_1,i_2,i_3,i_4} := E[Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}]$*. We use the abbreviation* $(Z_0,\ldots,Z_{p-1})^T = \underline{Z}(\Delta)$ *and denote by* $c_{j,i}$ *the* i*-th component of the vector* $\underline{b}^T\left(e^{A\Delta}\right)^j$*.*

Remark 4.5. *Of course Proposition 4.4 together with the delta-method immediately leads to asymptotic normality of the empirical autocorrelations,* $\widehat{\varrho}(h) = \widehat{\gamma}(h)/\widehat{\gamma}(0)$*,* $h = 1,2,\ldots$*. Comparing the expression (4.19) with the asymptotic covariance matrix of the empirical autocovariances of a discrete time linear process (cf. Brockwell and Davis (1991), Proposition 7.3.4) we see that both expressions consist of two summands. Although the second summands coincide, the first summands differ substantially. In Bartlett's formula for discrete time linear processes the asymptotic covariance matrix depends only on the autocorrelation function of the process. This convenient property fails to hold in the continuous time setting. Thus, in contrast with the expression in Brockwell and Davis (1991), Theorem 7.2.1, we obtain for the limiting covariance matrix, as* $n\to\infty$*,*

$$\sqrt{n}\Big(\widehat{\varrho}_Y(0)-\varrho_Y(0),\ldots,\widehat{\varrho}_Y(q)-\varrho_Y(q)\Big)^T \xrightarrow{\mathcal{D}} \mathcal{N}(0,W), \tag{4.20}$$

where the limiting covariance matrix $W = (w_{q_1,q_2})_{q_1,q_2=0,\dots,q} \in \mathbb{R}^{q+1,q+1}$ *reads*

$$\begin{aligned} & w_{q_1,q_2} \\ &= \frac{(\eta-3)\sigma^4}{\gamma_Y(0)^2} \int_0^1 \left(\sum_{k=-\infty}^{\infty} f(u+k)f(u+k+q_1) - \varrho_Y(q_1) f(u+k)^2 \right) \\ &\qquad \cdot \left(\sum_{l=-\infty}^{\infty} f(u+l)f(u+l+q_2) - \varrho_Y(q_2) f(u+l)^2 \right) du \\ &\quad + \sum_{k=1}^{\infty} [\varrho_Y(k+q_1) + \varrho_Y(k-q_1) - 2\varrho_Y(q_1)\varrho_Y(k)] \\ &\qquad \cdot [\varrho_Y(k+q_2) + \varrho_Y(k-q_2) - 2\varrho_Y(q_2)\varrho_Y(k)] \,. \end{aligned} \tag{4.21}$$

Here, as above, $f(t) = \underline{b}^T e^{At} \underline{e} \mathbf{1}_{[0,\infty)}(t)$*. For details we refer to Cohen and Lindner (2013) and, for a corresponding phenomenon for discrete-time linear processes observed at lower frequencies, to Niebuhr and Kreiss (2014).*

4.3 Bootstrap procedure

First we consider the simplest case, the CAR(1) or stationary Ornstein-Uhlenbeck process. In this case, both the observation and state equation (4.3) and (4.4) reduce to a single one-dimensional equation for $Y(t\Delta) = X(t\Delta)$. Namely

$$X((t+1)\Delta) \;=\; e^{-a\Delta} X(t\Delta) + Z((t+1)\Delta), \;\; t = 0, \dots, n-1. \tag{4.22}$$

Thus, every equidistantly discretely (fixed Δ-grid) sampled CAR(1) process is a first order autoregressive process with i.i.d. innovations. This, of course, is a very well-studied process in time series analysis. The autoregressive parameter $e^{-a\Delta}$ can be $\sqrt{n}$-consistently estimated using the Yule-Walker method, which immediately leads to a $\sqrt{n}$-consistent estimator of the continuous time parameter a via

$$\widehat{a} = \frac{-\log\left(\widehat{\gamma}(\Delta) \,/\, \widehat{\gamma}(0)\right)}{\Delta} \,. \tag{4.23}$$

Residual-based or wild bootstrap proposals are well understood for such cases and immediately lead to consistent bootstrap procedures for discretely observed Ornstein-Uhlenbeck processes.

Recall that our interest is to setup a bootstrap procedure which is able to consistently approximate distributions of statistics that depend on observations on a fixed

Δ-grid of the CAR process only. Without loss of generality, let us assume $\Delta = 1$. Consider as an important example empirical autocovariances $\widehat{\gamma}(h)$, cf. (4.16), or empirical autocorrelations. As Proposition 4.4 shows, the asymptotic variance of such quantities depends in a quite complicated way on properties of the underlying continuous time process which is quite difficult to estimate from discrete time observations. Thus, it appears that there is some room for a bootstrap procedure.

Extending the simple approach described above for CAR(1) processes to deal with CAR(p) processes with $p > 1$ presents serious difficulties. It is well-known (see e.g. Brockwell (1994), Huzii (2001)) that, from a second-order point of view, every discretely sampled CARMA(p, q) process can be represented as a stationary solution of ARMA(p, q') equations with $q' < p$. Brockwell and Lindner (2009) give a stronger result saying that the discretely sampled observations of a CARMA$(p,\ q)$ process satisfy autoregressive equations of order p with driving noise which is $(p-1)$-dependent. Since every $(p-1)$-dependent sequence has a moving average representation of order at most $(p-1)$ driven by white noise which is uncorrelated but not necessarily (except when $p = 1$) independent, our observations will satisfy an ARMA$(p, p-1)$ equation with innovations which are uncorrelated only. Thus a residual bootstrap as described in Kreiss and Franke (1992) using a standard ARMA$(p, p-1)$-model fit to the observations and a resampling via drawing with replacement from residuals from this fit will lead to consistent results only if statistics are considered whose asymptotic distribution depends only on second order properties, i.e. on the autocovariance structure of the observations. This is because a given ARMA model has autocovariances which are the same whether the driving noise is independent or simply uncorrelated. A simple example is the sample mean, $\overline{Y}$. Central limit results for $\overline{Y}$ can be established under quite general assumptions, which typically are satisfied for discretely observed CAR(p) processes. Since the asymptotic variance of $\overline{Y}$ depends only on the autocovariance function of Y, every bootstrap proposal which mimics the second order properties of the underlying observations will work asymptotically. However, in all cases with asymptotic distribution of a statistic of interest depending on properties that go beyond second order properties, such a simple ARMA-based residual bootstrap procedure for discretely observed CAR(p)-processes would fail! In Proposition 4.4 and Remark 4.5 we have seen that for empirical autocovariances and more interestingly, even for empirical autocorrelations, features of the process beyond second order properties show up in the asymptotic distribution. This fact directly implies that a standard residual based ARMA bootstrap does not work in

such situations.

The block bootstrap (cf. Künsch (1989) and Bühlmann and Künsch (1995)), which has been shown to work for rather general strictly stationary processes, is a possibility to overcome this problem. However we intend to follow in this chapter a different approach, which tries to take existing parametric structure as much as possible into account. Moreover block bootstrap techniques have to deal with quite delicate problems around a proper choice of the block length (e.g. Nordman et al. (2007)). Instead of dealing with block bootstrap methods we focus on an i.i.d. based bootstrap proposal influenced by the ideas of Kreiss and Franke (1992) and Paparoditis (1996). In order to be able to apply such a residual based bootstrap we make use of the vector autoregressive representation obtained from (4.12), namely

$$\underline{X}((t+1)\Delta) \;=\; e^{A\Delta}\underline{X}(t\Delta) + \underline{Z}((t+1)\Delta)\,. \tag{4.24}$$

In this vector autoregressive representation the driving white noise ($\underline{Z}(t+1)$ (cf. (4.13)) indeed is an i.i.d. noise sequence.

Our strategy now is to estimate the first $p-1$ derivatives $X_1(t\Delta), \ldots, X_{p-1}(t\Delta)$ of the CAR process $X_0(\cdot)$, which represent the back $p-1$ components of the vector $\underline{X}(t\Delta)$, and to use them to estimate the autoregressive parameter matrix $e^{A\Delta}$. Having done this we immediately are able to define estimated autoregessive residuals on which an asymptotically consistent residual based bootstrap can be set up.

To this end let us assume that we are able to observe some additional auxiliary high frequency data, but must point out that no full-time high frequency data is needed. More precisely we assume the following observation structure:

$$\begin{array}{cccc} Y_{1\Delta-(p-1)h}, & \ldots, & Y_{1\Delta-h}, & Y_{1\Delta}, \\ Y_{2\Delta-(p-1)h}, & \ldots, & Y_{2\Delta-h}, & Y_{2\Delta}, \\ \vdots & \vdots & \ldots & \vdots \\ Y_{n\Delta-(p-1)h}, & \ldots, & Y_{n\Delta-h}, & Y_{n\Delta}\,. \end{array} \tag{4.25}$$

In (4.25) $\Delta > 0$ still is the fixed grid size of our *main* observations $Y_{1\Delta}, \ldots, Y_{n\Delta}$. The auxiliary $p-1$ pre-observations are on a much finer high frequency grid of mesh size h, for which later will be assumed that $h \to 0$ as $n \to \infty$. This results in a local high-frequency aided low frequency sampling scheme (cf. Figure 4.1).

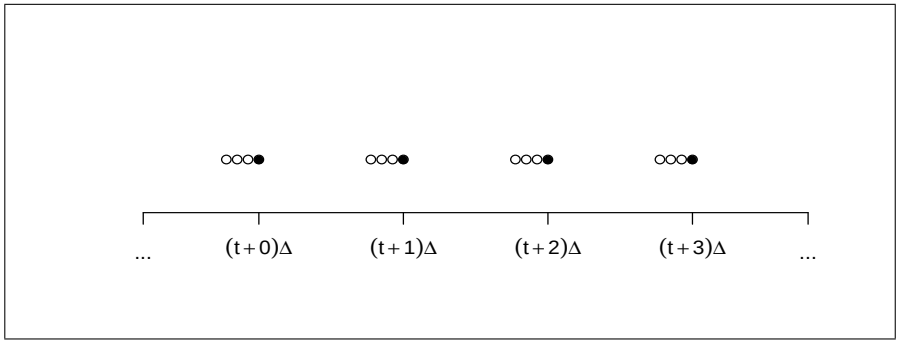

Figure 4.1: Local high-frequency aided low frequency sampling scheme.

Remark 4.6. *At first glance, the supposed data structure (cf. (4.25) or Figure 4.1) needed for the local high-frequency aided bootstrap proposal suggested seems somehow strange. The following two examples show the kind of situations our bootstrap proposal is applicable to. As a first example from financial econometrics one might be interested in fitting CAR(p) models on the basis of daily return data. Assume that* $\Delta = 1$ *stands for one-day length. For the application of our bootstrap proposal it is necessary to be able to additionally observe some more frequent intraday data, for example hourly, 30- or 15-minutes return values. This would lead to values* $h = 1/24$, $1/48$ *or* $1/96$, *respectively. Such higher frequency returns are available in many cases (e.g. for currency exchange rates and stock indices such as Dow Jones, S&P 500, FTSE 100, Nikkei or DAX). Alternatively we may have complete high frequency intraday return data available e.g. at equidistant intervals of 15 minutes. Then our bootstrap proposal allows us to mimic the distribution of autocovariances and autocorrelations for lags* $q \cdot \Delta$, $q = 0, 1, \ldots$ *on a coarser time grid (for example* $\Delta = 24$, *which corresponds to six hours).*

Remark 4.7. *For simplicity and easier understanding we assume all additional high-frequency observations to be on the same time grid of mesh size* h. *Nevertheless, it is worth emphasizing that the whole bootstrap proposal would yield the same asymptotic results if the pre-observations, let us say* $Y_{i\Delta - h_{i,j}}$, $i = 1, \ldots, n$, $j = 0, \ldots, p-1$ *had different spacings* $(h_{i,j} - h_{i,j-1})$, *where* $h_{i,0} = 0$. *For the procedure to work, it would suffice to assure*

$$\max_{i \in \{1,\ldots,n\},\, j \in \{1,\ldots,p-1\}} (h_{i,j} - h_{i,j-1}) \to_{n\to\infty} 0. \tag{4.26}$$

Based on the observations (4.25) the derivatives of the process are estimated consistently by Proposition 5.1 of Brockwell and Schlemm (2011) using iterated difference

quotients for the first $p-1$ derivatives. More precisely we define

$$\widehat{X_s}(t) := \frac{1}{h^s}\sum_{i=0}^{s} X_0(t-ih)(-1)^i\binom{s}{i}, \quad s=0,\dots,p-1\,. \tag{4.27}$$

If the driving Lévy process L is assumed to have finite second moments then Proposition 5.6 of Brockwell and Schlemm (2011) gives

$$\widehat{X_s}(t) = X_s(t) + \mathcal{O}_{\mathcal{P}}(h), \quad s=0,\dots,p-1. \tag{4.28}$$

The vectors

$$\widehat{\underline{X}}^T(t\Delta) = \left(X_0(t\Delta), \widehat{X_1}(t\Delta),\dots,\widehat{X_{p-1}}(t\Delta)\right)^T \tag{4.29}$$

are used to estimate the autoregressive parameter matrix, e.g. by the classical Yule-Walker equations. Thus

$$\widehat{e^{A\Delta}} = \widehat{\underline{\Gamma}}(\Delta)\widehat{\underline{\Gamma}}^{-1}(0)\,, \tag{4.30}$$

where

$$\widehat{\underline{\Gamma}}(\Delta) = \frac{1}{n}\sum_{t=1}^{n-\Delta}\left(\widehat{\underline{X}}((t+1)\Delta) - \widehat{\underline{X}}\right)\left(\widehat{\underline{X}}(t\Delta) - \widehat{\underline{X}}\right)^T. \tag{4.31}$$

Even though it would be sufficient for our purposes to have a consistent estimator of the autoregressive parameter matrix, inspection of the results of Brockwell and Schlemm (2011) leads to the following result:

Lemma 4.8. *Under Assumptions 2 and 3, and if $h = h(n) \to 0$ as $n \to \infty$, we obtain*

$$\widehat{e^{A\Delta}} = e^{A\Delta} + \mathcal{O}_{\mathcal{P}}\left(h + n^{-1/2}\right)\,. \tag{4.32}$$

Remark 4.9. *We emphasize that the autoregessive parameter matrix $e^{A\Delta}$ is estimated directly and not via an estimator of the matrix A itself composed with the matrix exponential function $e^{\cdot\Delta}$. Even if estimation of A were possible, the direct estimation of $e^{A\Delta}$ via Yule-Walker equations is much simpler in practice. Furthermore, it is well known that under very mild conditions the Yule-Walker estimate has eigenvalues with an absolute value less than one. This fact is a great advantage for the bootstrap procedure to be defined below. Moreover, except in the simple case when $p=1$, it is not immediately evident that for every fixed Δ the matrix exponential $e^{\cdot\Delta}$ can be inverted to produce from the estimator $\widehat{e^{A\Delta}}$ a uniquely defined estimator $\widehat{A}$ of the matrix A such that $\widehat{A}$ satisfies Assumption 2.*

The above considerations lead to the following bootstrap algorithm which is used to generate pseudo-observations $Y^*(\Delta), Y^*(2\,\Delta), \ldots, Y^*(n\,\Delta)$ of the continuous-time CAR(p) process $(Y(t) : t \geq 0)$.

Step 1:
Let $\widehat{e^{A\Delta}}$ denote a consistent estimator of $e^{A\Delta}$. Then, obtain estimated residuals from

$$\underline{\widehat{Z}}(t\Delta) = \underline{X}(t\Delta) - \widehat{e^{A\Delta}}\underline{X}((t-1)\Delta),\ t = 1, \ldots n\,. \tag{4.33}$$

Step 2:
Generate bootstrap residuals $(\underline{Z}^*(t\,\Delta))$ via drawing with replacement from the centered esimated innovations $\underline{\widehat{Z}^c}(\Delta), \ldots, \underline{\widehat{Z}^c}(n\,\Delta)$, where the centering is done by the usual way via $\underline{\widehat{Z}^c}(t\,\Delta) = \underline{\widehat{Z}}(t\,\Delta) - 1/n\,\sum_{j=1}^{n}\underline{\widehat{Z}}(j\,\Delta)$.

Step 3:
Obtain pseudo-observations $\underline{X}^*(t\Delta)$, $t = 1, \ldots, n$, of the vector autoregressive process from

$$\underline{X}^*(t\Delta) = \widehat{e^{A\Delta}}\underline{X}^*((t-1)\Delta) + \underline{Z}^*(t\,\Delta)\,. \tag{4.34}$$

Step 4:
Finally obtain pseudo-observations $Y^*(\Delta), Y^*(2\,\Delta), \ldots, Y^*(n\,\Delta)$ according to

$$Y^*(t\,\Delta) = (1, 0, \ldots, 0)\,\underline{X}^*(t\Delta),\ t = 1 \ldots, n\,. \tag{4.35}$$

Exactly as for the vector autoregressive process (4.12) we obtain for the bootstrapped vector autoregression $(\underline{X}^*(t\Delta))$ a moving average representation of form

$$\underline{X}^*((t+1)\Delta) \;=\; \sum_{j=0}^{\infty} \widehat{e^{A\Delta j}}\underline{Z}^*((t+1-j)\Delta)\,. \tag{4.36}$$

This directly leads to a bootstrap analogue of (4.15) which is

$$Y^*((t+1)\Delta) = \sum_{j=0}^{\infty}\sum_{i=0}^{p-1} \widehat{c}_{j,i} Z_i^*((t+1-j)\Delta)\,, \tag{4.37}$$

where $(\widehat{c}_{j,0}, \ldots, \widehat{c}_{j,p-1}) = \underline{\widehat{c}}_j^T := \underline{b}^T \widehat{e^{A\Delta j}}$, $j \in \mathbb{N}_0$.

In Section 4.6 we prove the following result, which states that our bootstrap proposal works asymptotically for statistics depending smoothly on autocovariances or autocorrelations.

Theorem 4.10. *Let Y be a CAR(p) process and let Assumptions 2 and 3 be satisfied. Further assume the local high-frequency aided sampling scheme (4.25) with h satisfying $h = h(n) \to 0$ as $n \to \infty$ and let $Y^*(t\,\Delta)$ be a bootstrap process generated as described above. Then we have in probability as $n \to \infty$:*

(i) For each $q_1, q_2 \in \mathbb{N}_0$ and $\widehat{\gamma}^*(q_j\Delta) = n^{-1}\sum_{t=1}^{n-q_j} Y^*(t\Delta)Y^*((t+q_j)\Delta)$, $j = 1, 2$, it holds

$$\lim_{n\to\infty} n\ Cov(\widehat{\gamma}^*(q_1\Delta), \widehat{\gamma}^*(q_2\Delta)) \to v_{q_1,q_2,\Delta}, \tag{4.38}$$

where $v_{q_1,q_2,\Delta}$ is defined as in Proposition 4.4.

(ii) Further for each $q \in \mathbb{N}_0$, it holds

$$\sqrt{n}\Big(\widehat{\gamma}^*(0) - \gamma_{Y^*}(0), \ldots, \widehat{\gamma}^*(q\Delta) - \gamma_{Y^*}(q\Delta)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)\,, \tag{4.39}$$

where $\gamma_{Y^*}(\cdot)$ denotes the autocovariance function of the bootstrap sequence $(Y^*(t\Delta))$ (cf. (4.35)) and $V = (v_{q_1,q_2,\Delta})_{q_1,q_2=0,\ldots,q}$.

(iii) Moreover, it holds

$$\sqrt{n}\Big(\widehat{\varrho}^*(0) - \varrho_{Y^*}(0), \ldots, \widehat{\varrho}^*(q\Delta) - \varrho_{Y^*}(q\Delta)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W)\,, \tag{4.40}$$

where $\varrho_{Y^*}(\cdot)$ denotes the autocorrelation function of the bootstrap sequence $(Y^*(t\Delta))$ and W is as given in Remark 4.5.

Remark 4.11. *Since the bootstrap procedure proposed in Section 4.3 mimics the true underlying vector autoregressive process (including the distribution of the errors) it can be expected that the validity of our bootstrap proposal goes far beyond statistics which are smooth functionals of empirical autocovariances and autocorrelations. Especially for integrated periodograms (cf. Dahlhaus (1985)), nonparametric spectral density estimation and the general class of estimators*

$$T_n = f\left(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g\left(Y(t\Delta), \ldots, Y((t+m-1)\Delta)\right)\right), \tag{4.41}$$

discussed in Künsch (1989), cf. Example 2.2; for $g : \mathbb{R}^m \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$, our proposal will lead to a consistent approximation of the distribution of the corresponding statistics. For the latter class of statistics, Bühlmann (1997) proved validity of the AR-sieve bootstrap under the main assumption of an invertible linear process in discrete time with i.i.d. innovations for the underlying process. One should keep

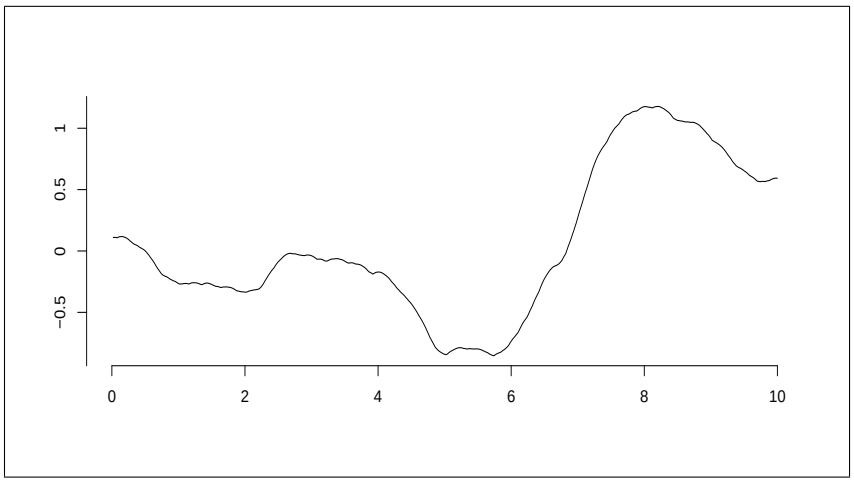

Figure 4.2: Typical realization of a CAR(2) process.

in mind that the proposed bootstrap procedure is aimed at approximating distributions of statistics that can be written as functionals of discretely observed data (fixed Δ-grid) from a CAR(p) process of known order p. The additional high-frequency pre-observations preceding each time point $t\Delta$ are only auxiliary values in order to approximate derivatives of the underlying process at the time points $t\Delta$.

4.4 A simulation study

In this section we present the results of a simulation study for CAR(2) processes. We simulated a CAR(2) process with parameters $a_1 = -1.0525$ and $a_2 = -1.5$. A Wiener process with variance 1 was used as the underlying driving Lévy process. Figure 4.2 shows a typical realization of such a process. Note that the smooth appearing of the sample path is quite expected because CAR(2) processes are differentiable. We set $n = 150$ and $\Delta = 1$ and investigated the finite sample distribution of the first order autocorrelation

$$\sqrt{n}\,(\widehat{\varrho}(1) - \varrho(1))\ , \tag{4.42}$$

based on observations as given in (4.25), and the ability of the proposed bootstrap proposal to approximate this distribution. We simulated (4.42) 1,500 times to get an appropriate approximation of the finite sample distribution.

The histograms in red color in Figure 4.3 show this simulated finite sample distribution of (4.42). The histograms in grey are bootstrap distributions showing aver-

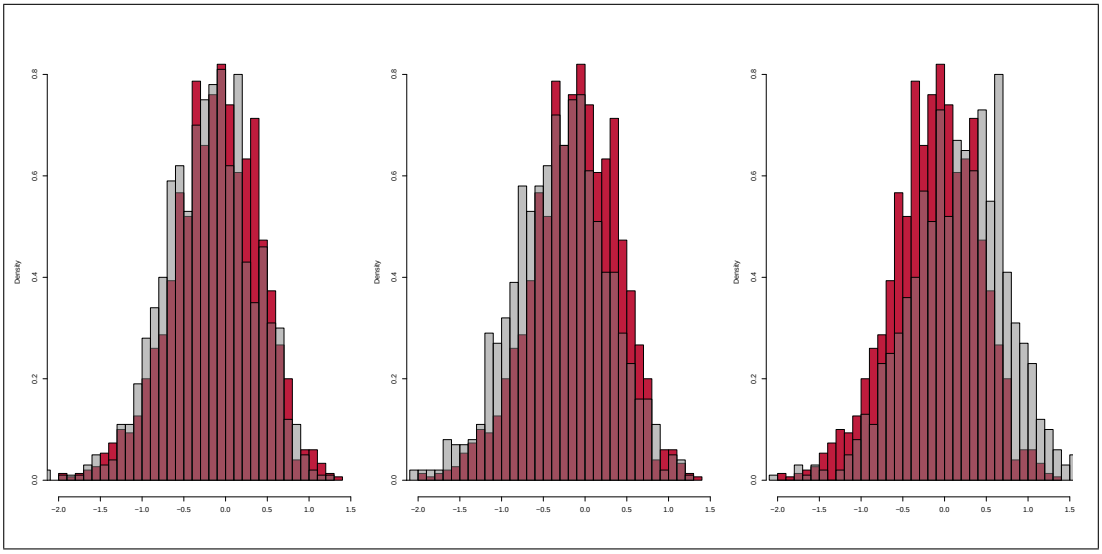

Figure 4.3: Average bootstrap performance (lower quartile, median and upper quartile distance). True finite sample distribution (red) and bootstrap approximations (light grey) of $\mathcal{L}(\sqrt{n}(\widehat{\varrho}(1) - \varrho(1)))$. Overlapping area of both histograms is in dark grey.

age performance. Showing average performance in this context means that we have simulated 1,000 bootstrap distributions and have calculated their distances to the true distribution (histogram in red color). The grey histogram plots now represent bootstrap distributions belonging to the lower quartile, the median and the upper quartile of distances, respectively. Even if we have not incorporated the limiting normal distribution in the simulation, because the limiting variance (cf. Proposition 4.4) is quite difficult to compute, it appears that the true distribution shows a significant skewness. Thus, it can be expected that the limiting normal distribution will possess certain approximation errors. Further, Figure 4.4 shows boxplots of generated bootstrap 5%, 95% and 99% quantiles. The added red lines represent the corresponding true quantiles obtained by simulation.

4.5 Conclusions

We have proposed a bootstrap procedure which is applicable to discrete time (fixed Δ-grid) observations from continuous-time autoregressive processes. Starting from the Ornstein-Uhlenbeck process as the simplest CAR process, for which a consistent bootstrap procedure easily can be defined, we have seen that the situation becomes much more complicated for samples from CAR processes of higher order. By using some auxiliary high frequency pre-observations preceding every discrete

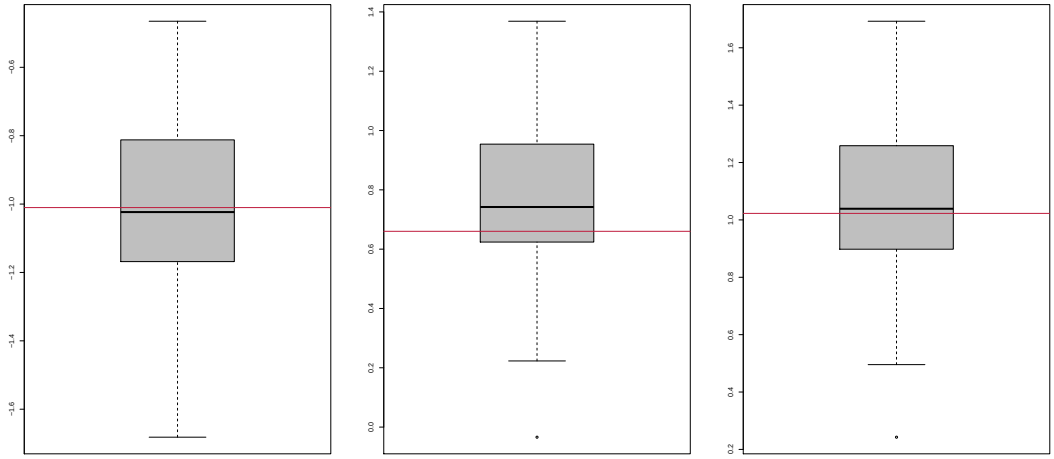

Figure 4.4: Boxplots of 5%, 95% and 99% bootstrap quantiles of $\mathcal{L}(\sqrt{n}(\widehat{\varrho}(1)-\varrho(1)))$ (left to right) with true quantiles (in red).

low frequency time point we make use of the fact that the process together with its first $p-1$ derivatives can be written as a vector autoregressive process of order one and, most important for the bootstrap, with i.i.d. innovations. On this basis a bootstrap procedure has been proposed and the asymptotic validity has been shown for empirical autocovariances and empirical autocorrelations. It has been pointed out, that the normal approximation for the distributions of empirical autocovariances and empirical autocorrelations differs from that for linear time series in discrete time. Even the asymptotic variance of such limiting normal distributions can hardly be estimated from low-frequency data. A small simulation study has shown that the proposed bootstrap proposal works appropriately.

4.6 Proofs and auxiliary results

Proof of Proposition 4.4:
Since all the main arguments for a proof of Proposition 4.4 have been given in Cohen and Lindner (2013) we restrict ourselves to verify the representation (4.19) which differs from the representation given in Cohen and Lindner (2013) but is needed for the proof of our main result. We make heavy use of (4.15) in the following. Obviously $E[Y(t\Delta)] = 0$. Further we obtain

$$\gamma_Y(h\Delta) = E[Y(t\Delta)Y((t+h)\Delta)] = \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1} c_{j,i_1}c_{j+h,i_2}E[Z_{i_1}Z_{i_2}]\,. \tag{4.43}$$

With the notation $\kappa^4_{i_1,i_2,i_3,i_4} = E[Z_{i_1}Z_{i_2}Z_{i_3}Z_{i_4}]$ we can directly compute

$$\begin{aligned}
&E[Y(t\Delta)Y((t+q_1)\Delta)Y((t+h+q_1)\Delta)Y((t+h+q_1+q_2)\Delta)] \qquad (4.44)\\
&= \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} c_{j,i_1}c_{j+q_1,i_2}c_{j+q_1+h,i_3}c_{j+q_1+h+q_2,i_4}\\
&\quad\cdot\Big(\kappa^4_{i_1,i_2,i_3,i_4} - E[Z_{i_1}Z_{i_2}]E[Z_{i_3}Z_{i_4}] - E[Z_{i_1}Z_{i_3}]E[Z_{i_2}Z_{i_4}] - E[Z_{i_1}Z_{i_4}]E[Z_{i_2}Z_{i_3}]\Big)\\
&\quad+\gamma_Y(q_1\Delta)\gamma_Y(q_2\Delta) + \gamma_Y((q_1+h)\Delta)\gamma_Y((h+q_2)\Delta)\\
&\quad+\gamma_Y((q_1+h+q_2)\Delta)\gamma_Y(h\Delta).
\end{aligned}$$

This last representation corresponds to equation (3.5) in Cohen and Lindner (2013). The next step is to compute the asymptotic behaviour of $n\ Cov(\widehat{\gamma}_Y(q_1\Delta), \widehat{\gamma}_Y(q_2\Delta))$ for general integers q_1, q_2. By the same strategy as in the proof of Lemma 2.9, using the unbiased (and asymptotically equivalent) estimate

$$\widetilde{\gamma}_Y(p\Delta) = n^{-1}\sum_{t=1}^{n} Y(t\Delta)Y((t+p)\Delta) \qquad (4.45)$$

instead of $\widehat{\gamma}_Y(\cdot)$ within the computations, one easily observes that

$$\begin{aligned}
&n\ Cov(\widehat{\gamma}_Y(q_1\Delta), \widehat{\gamma}_Y(q_2\Delta)) \qquad (4.46)\\
&= n\left[E[\widetilde{\gamma}_Y(q_1\Delta)\widetilde{\gamma}_Y(q_2\Delta)] - E[\widetilde{\gamma}_Y(q_1\Delta)]E[\widetilde{\gamma}_Y(q_2\Delta)]\right] + o(1)\\
&= \sum_{r=-(n-1)}^{n-1}\left(1-\frac{|r|}{n}\right)T_r + o(1)\,,
\end{aligned}$$

where

$$\begin{aligned}
T_r \;=\; &\sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} c_{j,i_1}c_{j+q_1,i_2}c_{j+r,i_3}c_{j+r+q_2,i_4} \qquad (4.47)\\
&\cdot\Big(\kappa^4_{i_1,i_2,i_3,i_4} - E[Z_{i_1}Z_{i_2}]E[Z_{i_3}Z_{i_4}] - E[Z_{i_1}Z_{i_3}]E[Z_{i_2}Z_{i_4}] - E[Z_{i_1}Z_{i_4}]E[Z_{i_2}Z_{i_3}]\Big)\\
&+\gamma_Y(r\Delta)\gamma_Y((q_2+r-q_1)\Delta) + \gamma_Y((r+q_2)\Delta)\gamma_Y((r-q_1)\Delta)\,.
\end{aligned}$$

Then, as $n\to\infty$, (4.46) yields (4.19) by the weak law of large numbers and the summability of the coefficients:

$$\begin{aligned}
&n\ Cov(\widetilde{\gamma}_Y(q_1\Delta), \widetilde{\gamma}_Y(q_2\Delta)) \qquad (4.48)\\
&\overset{n\to\infty}{\longrightarrow} \sum_{r=-\infty}^{\infty} T_r\\
&= \sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1}\sum_{j=0}^{\infty} c_{j,i_1}c_{j+q_1,i_2}\sum_{r=0}^{\infty} c_{j+r,i_3}c_{j+r+q_2,i_4}
\end{aligned}$$

$$\cdot(\kappa^4_{i_1,i_2,i_3,i_4} - E[Z_{i_1}Z_{i_2}]E[Z_{i_3}Z_{i_4}] - E[Z_{i_1}Z_{i_3}]E[Z_{i_2}Z_{i_4}] - E[Z_{i_1}Z_{i_4}]E[Z_{i_2}Z_{i_3}])$$
$$+ \sum_{r=-\infty}^{\infty} \gamma_Y(r\Delta)\gamma_Y((q_2+r-q_1)\Delta) + \gamma_Y((r+q_2)\Delta)\gamma_Y((r-q_1)\Delta)$$

This concludes the proof that (4.19) is found to be another representation of equation (3.4) in Cohen and Lindner (2013). □

Proof of Theorem 4.10:
The computation of the asymptotic covariance matrix is exactly as in the proof of Proposition 4.4. We obtain

$$\begin{aligned}
&n\ Cov(\widehat{\gamma}_{Y^*}(q_1\Delta), \widehat{\gamma}_{Y^*}(q_2\Delta)) \qquad (4.49)\\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\Bigg[\sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1}\widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2}\widehat{c}_{j+s-t,i_3}\widehat{c}_{j+s-t+q_2,i_4}\\
&\quad\cdot\Big(\kappa^{*4}_{i_1,i_2,i_3,i_4} - E^*[Z^*_{i_1}Z^*_{i_2}]E^*[Z^*_{i_3}Z^*_{i_4}]\\
&\qquad\qquad - E^*[Z^*_{i_1}Z^*_{i_3}]E^*[Z^*_{i_2}Z^*_{i_4}] - E^*[Z^*_{i_1}Z^*_{i_4}]E^*[Z^*_{i_2}Z^*_{i_3}]\Big)\Bigg]\\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}[\gamma_{Y^*}((s-t)\Delta)\gamma_{Y^*}((q_2+s-t-q_1)\Delta)\\
&\qquad + \gamma_{Y^*}((s-t+q_2)\Delta)\gamma_{Y^*}((s-t-q_1)\Delta)] + o_P(1)\\
&= \sum_{r=-(n-1)}^{n-1}\left(1-\frac{|r|}{n}\right)T^*_r + o_P(1)\,.
\end{aligned}$$

Here $\kappa^{4*}_{i_1,i_2,i_3,i_4} = E^*[Z^*_{i_1}Z^*_{i_2}Z^*_{i_3}Z^*_{i_4}]$ and

$$\begin{aligned}
T^*_r \;=\; &\sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1}\widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2}\widehat{c}_{j+r,i_3}\widehat{c}_{j+r+q_2,i_4} \qquad (4.50)\\
&\cdot\Big(\kappa_4{}^*_{i_1,i_2,i_3,i_4} - E^*[Z^*_{i_1}Z^*_{i_2}]E^*[Z^*_{i_3}Z^*_{i_4}] - E^*[Z^*_{i_1}Z^*_{i_3}]E^*[Z^*_{i_2}Z^*_{i_4}]\\
&\qquad - E^*[Z^*_{i_1}Z^*_{i_4}]E^*[Z^*_{i_2}Z^*_{i_3}]\Big)\\
&+\gamma_{Y^*}(r\Delta)\gamma_{Y^*}((q_2+r-q_1)\Delta) + \gamma_{Y^*}((r+q_2)\Delta)\gamma_{Y^*}((r-q_1)\Delta)\,.
\end{aligned}$$

Because of Assumption 2 the parameter matrix $e^{A\Delta}$ only has eigenvalues within the unit circle. The same holds for $\widehat{e^{A\Delta}}$ reasoned by well known properties of the Yule-Walker method (cf. Remark 4.9). Thus the matrix polynomials

$$I_p - e^{A\Delta}z \quad \text{and} \quad I_p - \widehat{e^{A\Delta}}z \qquad (4.51)$$

only have roots outside the closed unit circle, and

$$\left(I_p - e^{A\Delta}z\right)^{-1} \quad \text{and} \quad \left(I_p - \widehat{e^{A\Delta}}z\right)^{-1} \tag{4.52}$$

can be written as power series

$$\sum_{j=0}^{\infty}\left(e^{A\Delta}\right)^j z^j \quad \text{and} \quad \sum_{j=0}^{\infty}\left(\widehat{e^{A\Delta}}\right)^j z^j, \tag{4.53}$$

respectively, for all $|z| \leq 1+\delta$ and some $\delta > 0$. Using a multidimensional version of Cauchy's inequality for holomorphic functions we obtain

$$\sup_{j\in\mathbb{N}_0}(1+\delta)^j \left\|\left(\widehat{e^{A\Delta}}\right)^j - \left(e^{A\Delta}\right)^j\right\| = \mathcal{O}_P(h+n^{-1/2})\,, \tag{4.54}$$

for some $\delta > 0$ (cf. Jentsch and Kreiss (2010)) equation (7.7) for the case $p = 1$), where $\|\cdot\|$ denotes the Euclidean matrix norm. Because $\underline{c}_j$, $\widehat{\underline{c}}_j$ is just the first row of the matrix $\left(e^{A\Delta}\right)^j$, $\left(\widehat{e^{A\Delta}}\right)^j$, respectively, (4.54) immediately leads to

$$\sup_{j\in\mathbb{N}_0, i=0,\ldots,p-1} r^j\ |\widehat{c}_{j,i} - c_{j,i}| = \mathcal{O}_P(h+n^{-1/2})\,. \tag{4.55}$$

Equation (4.55) together with consistency of ${\kappa^4}^*_{i_1,i_2,i_3,i_4}$ and $E^*[Z_i^*Z_j^*]$ for ${\kappa^4}_{i_1,i_2,i_3,i_4}$ and $E[Z_iZ_j]$ respectively (both are immediate consequences of the weak law of large numbers) as well as the summability of the coefficients $c_{j,i}$ and $\widehat{c}_{j,i}$ now leads by a direct but tedious computation to the result

$$\sum_{r=-(n-1)}^{n-1} |T_r^* - T_r| = o_P(1)\,, \tag{4.56}$$

which means by (4.49) that

$$n\ Cov(\widehat{\gamma}_{Y^*}(q_1\Delta), \widehat{\gamma}_{Y^*}(q_2\Delta)) \to v_{q_1\Delta,q_2\Delta}\,. \tag{4.57}$$

The explicit computation is as follows. Denote

$$\begin{aligned} C_4^*(n) \ :=\ & {\kappa_4}^*_{i_1,i_2,i_3,i_4} - E^*[Z_{i_1}^*Z_{i_2}^*]E^*[Z_{i_3}^*Z_{i_4}^*] - E^*[Z_{i_1}^*Z_{i_3}^*]E^*[Z_{i_2}^*Z_{i_4}^*] \\ & - E^*[Z_{i_1}^*Z_{i_4}^*]E^*[Z_{i_2}^*Z_{i_3}^*] \end{aligned} \tag{4.58}$$

and

$$\begin{aligned} C_4 \ :=\ & \kappa_{4i_1,i_2,i_3,i_4} - E[Z_{i_1}Z_{i_2}]E[Z_{i_3}Z_{i_4}] - E[Z_{i_1}Z_{i_3}]E[Z_{i_2}Z_{i_4}] \\ & - E[Z_{i_1}Z_{i_4}]E[Z_{i_2}Z_{i_3}], \end{aligned} \tag{4.59}$$

and further,

$$\begin{aligned} C_2^*(n) &:= \gamma_{Y^*}(r\Delta)\gamma_{Y^*}((q_2+r-q_1)\Delta) \\ &\quad +\gamma_{Y^*}((r+q_2)\Delta)\gamma_{Y^*}((r-q_1)\Delta) \end{aligned} \tag{4.60}$$

and

$$\begin{aligned} C_2 &:= \gamma_{Y}(r\Delta)\gamma_{Y}((q_2+r-q_1)\Delta) \\ &\quad +\gamma_{Y}((r+q_2)\Delta)\gamma_{Y}((r-q_1)\Delta)\,. \end{aligned} \tag{4.61}$$

Then, compute

$$\begin{aligned} &\sum_{r=-(n-1)}^{n-1} |T_r^* - T_r| \qquad (4.62)\\ &= \sum_{r=-(n-1)}^{n-1} \Bigg| \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} \widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2}\widehat{c}_{j+r,i_3}\widehat{c}_{j+r+q_2,i_4} C_4^*(n) \\ &\qquad\qquad - \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} c_{j,i_1}c_{j+q_1,i_2}c_{j+r,i_3}c_{j+r+q_2,i_4} C_4 \\ &\qquad + (C_2^*(n) - C_2) \Bigg| \\ &\le \sum_{r=-(n-1)}^{n-1} \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} \Big| \widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2}\widehat{c}_{j+r,i_3}\widehat{c}_{j+r+q_2,i_4} C_4^*(n) \\ &\qquad\qquad - c_{j,i_1}c_{j+q_1,i_2}c_{j+r,i_3}c_{j+r+q_2,i_4} C_4 \Big| \\ &\qquad + \Big| C_2^*(n) - C_2 \Big| \\ &= \sum_{r=-(n-1)}^{n-1} \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} \\ &\qquad \Big| \widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2}(\widehat{c}_{j+r,i_3} - c_{j+r,i_3})(\widehat{c}_{j+r+q_2,i_4} - c_{j+r+q_2,i_4}) C_4^*(n) \\ &\qquad + \widehat{c}_{j,i_1}\widehat{c}_{j+q_1,i_2} c_{j+r,i_3}c_{j+r+q_2,i_4} C_4^*(n) \\ &\qquad - c_{j,i_1}c_{j+q_1,i_2}c_{j+r,i_3}c_{j+r+q_2,i_4} C_4 \Big| \\ &\qquad + \Big| C_2^*(n) - C_2 \Big| \\ &= \sum_{r=-(n-1)}^{n-1} \sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1} \\ &\quad \Big| (\widehat{c}_{j,i_1} - c_{j,i_1})(\widehat{c}_{j+q_1,i_2} - c_{j+q_1,i_2}) \end{aligned}$$

$$
\begin{aligned}
&\cdot(\widehat{c}_{j+r,i_3}-c_{j+r,i_3})(\widehat{c}_{j+r+q_2,i_4}-c_{j+r+q_2,i_4})C_4^*(n)\\
&+c_{j,i_1}c_{j+q_1,i_2}(\widehat{c}_{j+r,i_3}-c_{j+r,i_3})(\widehat{c}_{j+r+q_2,i_4}-c_{j+r+q_2,i_4})C_4^*(n)\\
&+(\widehat{c}_{j,i_1}-c_{j,i_1})(\widehat{c}_{j+q_1,i_2}-c_{j+q_1,i_2})c_{j+r,i_3}c_{j+r+q_2,i_4}C_4^*(n)\\
&+c_{j,i_1}c_{j+q_1,i_2}c_{j+r,i_3}c_{j+r+q_2,i_4}(C_4^*(n)-C_4)\\
&+\Big|C_2^*(n)-C_2\Big|\\
\le&\sum_{r=-(n-1)}^{n-1}\sum_{j=0}^{\infty}\sum_{i_1=0}^{p-1}\sum_{i_2=0}^{p-1}\sum_{i_3=0}^{p-1}\sum_{i_4=0}^{p-1}\\
&|\widehat{c}_{j,i_1}-c_{j,i_1}|\cdot|\widehat{c}_{j+q_1,i_2}-c_{j+q_1,i_2}|\\
&\qquad\cdot|\widehat{c}_{j+r,i_3}-c_{j+r,i_3}|\cdot|\widehat{c}_{j+r+q_2,i_4}-c_{j+r+q_2,i_4}|\cdot|C_4^*(n)|\\
&+|c_{j,i_1}|\cdot|c_{j+q_1,i_2}|\cdot|\widehat{c}_{j+r,i_3}-c_{j+r,i_3}|\cdot|\widehat{c}_{j+r+q_2,i_4}-c_{j+r+q_2,i_4}|\cdot|C_4^*(n)|\\
&+|\widehat{c}_{j,i_1}-c_{j,i_1}|\cdot|\widehat{c}_{j+q_1,i_2}-c_{j+q_1,i_2}|\cdot|c_{j+r,i_3}|\cdot|c_{j+r+q_2,i_4}|\cdot|C_4^*(n)|\\
&+|c_{j,i_1}|\cdot|c_{j+q_1,i_2}|\cdot|c_{j+r,i_3}|\cdot|c_{j+r+q_2,i_4}|\cdot|C_4^*(n)-C_4|\\
&+|C_2^*(n)-C_2|.
\end{aligned}
$$

By the consistency of ${\kappa^4}^*_{i_1,i_2,i_3,i_4}$ and $E^*[Z_i^*Z_j^*]$ for ${\kappa^4}_{i_1,i_2,i_3,i_4}$ and $E[Z_iZ_j]$ we immediately have the consistency of $C_4^*(n)$ and $C_2^*(n)$ for C_4 and C_2. As $n\to\infty$, this then cancels the last two expresssions above. The remainder sums all converge to zero by the summability of the coefficients, the boundedness of $C_4^*(n)$ and especially (4.55). Altogether this proofs (4.56) and leads to (4.57). This is part (i) of Theorem 4.10.

For a proof of part (ii) of Theorem 4.10 we make use of Brockwell and Davis (1991), Proposition 6.3.9. Recall (4.37) and define $Y_M^*((t+1)\Delta)=\sum_{j=0}^M \widehat{c}_{j,i}Z_i^*((t+1-j)\Delta), M\in\mathbb{N}$. This sequence is M-dependent and a slight extension to triangular arrays (cf. Lemma 4.12 below) of the central limit theorem for M-dependent sequences stated in Brockwell and Davis (1991), Theorem 6.4.2, leads us to the asymptotic normality

$$
\sqrt{n}\Big(\widehat{\gamma}_M^*(0)-\gamma_{Y^*,M}(0),\ldots,\widehat{\gamma}_M^*(q\Delta)-\gamma_{Y^*,M}(q\Delta)\Big)\xrightarrow{\mathcal{D}}\mathcal{N}(0,V_M)\,, \quad (4.63)
$$

where $\widehat{\gamma}_M^*(h)$ and $\gamma_{Y^*,M}(h)$ are defined as $\widehat{\gamma}^*(h)$ and $\gamma_{Y^*}(h)$ with Y^* replaced by Y_M^*. V_M is defined as in (4.19) with ∞ replaced by M. Since $V_M\to V$ as $M\to\infty$ and for every $\varepsilon>0$ and every $h\in\mathbb{N}_0$

$$
\begin{aligned}
\lim_{M\to\infty}\limsup_{n\to\infty}P\Big\{\Big|&\sqrt{n}(\widehat{\gamma}^*(h\Delta)-\gamma_{Y^*}(h\Delta)) \qquad (4.64)\\
&-\sqrt{n}(\widehat{\gamma}_M^*(h\Delta)-\gamma_{Y^*,M}(h\Delta))\Big|>\varepsilon\Big\}=0
\end{aligned}
$$

the above mentioned result (i.e., Brockwell and Davis (1991), Proposition 6.3.9) yields part (ii).

Finally we obtain part (iii) by the usual delta-method from (ii) since autocorrelations are smooth functions of autocovariances. This concludes the proof of Theorem 4.10. □

In the proof of Theorem 4.10 we have made use of the following central limit theorem for triangular arrays of M-dependent sequences. We note that the truncated bootstrap process $Y_M^*(t\Delta)$ is indeed a triangular array of M-dependent random variables since with increasing n the parameters $\widehat{c}_{j,i}$ as well as the distribution of $Z_i^*(j\Delta)$ vary.

Lemma 4.12. *Suppose that for each $n \in \mathbb{N}$, real-valued, centered and M-dependent ($M \in \mathbb{N}$) random variables $\{U_{t,n} : t = 1, \ldots, n\}$ are given and make the following assumptions.*

(i) For $h \in \mathbb{N}_0$ we have $E\left(U_{t+h,n}U_{t,n}\right) \to_{n\to\infty} c(h)$, $h \in \mathbb{N}_0$, where the function c fulfills $c(0) + 2\sum_{k=1}^{m} c(h) = \tau^2 > 0$.

(ii) $\frac{1}{n^{1+\delta}} \sum_{t=1}^{n} E\left|U_{t,n}\right|^{2(1+\delta)} \to_{n\to\infty} 0$ for some $\delta > 0$.

Then we have

$$\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_{t,n}\right) \;=\; \tau^2 \tag{4.65}$$

and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_{t,n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2)\,. \tag{4.66}$$

Proof of Lemma 4.12:
For a proof of this central limit result we refer to Kreiss (1997), Lemma 8.4 in combination with the ensuing remark therein. □

5 | Autoregressive-aided block bootstrap

Based on : Niebuhr, T., Kreiß, J.-P. and Paparoditis, E. (2014):
Autoregressive-aided block bootstrap.
In preparation.

Abstract. We introduce a new block-based bootstrap procedure for general stationary sequences, called the autoregressive-aided block bootstrap, which generalizes both the moving block bootstrap as well as the residual-based bootstrap. The autoregressive-aided block bootstrap consists of two main steps, namely an autoregressive model fit and an ensuing (moving) block bootstrap. The parametric model-fit prewhitens the time series such that the dependence structure of the remaining residuals is simplified and closer to a white noise sequence. This allows the moving block bootstrap to capture more appropriate the interdependencies and to result in better performances compared to the ordinary moving block bootstrap. Furthermore, the proposal is shown to be robust against misestimated model orders which is an substantial advantage over the residual-based bootstrap. Indeed, we neither require nor attempt a reduction to independent and identically distributed values. The new procedure is shown to be valid for the large class of functions of generalized means. Comments on second-order correctness are given. Simulation studies on the finite-sample performances for several scenarios together with some short real-world applications will conclude this chapter.

5.1 Introduction

The block-based bootstrap for time series has been actively investigated since Künsch (1989) and Liu and Singh (1992) independently introduced the moving block bootstrap (MBB). The method has been investigated in quite some detail (e.g. Politis and Romano (1992), Shao and Yu (1993), Naik-Nimbalkar and Rajarshi (1994), Bühlmann (1994), Bühlmann and Künsch (1995), among others). The MBB is an important extension of Efron's bootstrap (Efron (1979)) allowing for dependent observations and does not assume any underlying parametric model. Parametric approaches reduce the time series to independent and identically distributed (i.i.d.) innovations to which Efron's bootstrap is applied. This leads to theoretically valid bootstrap approaches (e.g. Freedman (1984)), which crucially depend on the model adaptation. The advantage of the MBB is its generality and the fact that it "can be applied to complicated situations where parametric modeling and/or theoretical analysis is hopeless" (Efron (1982)). Based on the idea of resampling blocks several variants of the MBB have been stated; among them the non-overlapping block bootstrap (Carlstein (1986)), the circular block bootstrap (Politis and Romano (1992)), the matched block bootstrap (Carlstein et al. (1998)), and the tapered block bootstrap (Paparoditis and Politis (2001)). For a detailed discussion we refer the reader to Lahiri (2003).

The main idea of the block bootstrap is to cut the observed time series into blocks of some length l and then to resample the blocks instead of single observations. The procedure thus relies on producing a compromise between preserving the dependence structure of the original data within each block and corrupting it by supposing that the data is independent. This clarifies why the block bootstrap succeeds. It correctly mimics the l-dimensional distribution of the original process. Then, since the block size increases as the number of observations increases (at an appropriate rate), the block bootstrap succeeds in mimicking the entire dependence structure of the process. Clearly, on the one hand, this approximation is best if the dependence is weak and the blocks are as long as possible. On the other hand, the more distinct values are resampled the better the goodnis of the estimated distribution, and this points towards short blocks. For that reason, all block based methods have in common that they are sensitive to the block length l. When inappropriately chosen, the obtained results may perform unsatisfyingly bad. The optimal block size usually will depend on context (Hall et al. (1995)), nevertheless several studies investigating optimal

rates of block lengths were done (e.g. Lahiri (1999), Bühlmann and Künsch (1999), Lahiri et al. (2007) or more recent Nordman and Lahiri (2014)).

In this chapter we propose a new block-based bootstrap approach, which we call the autoregressive-aided block bootstrap (ARAB), that is generally applicable to any stationary, weakly dependent time series as the MBB. The basic idea of the ARAB is to *prewhiten* the time series by using a model fit, first, and then to apply the block bootstrap to the resulting (pseudo-)residuals. This approach was previously discussed in Davison and Hinkley (1997), Section 8.2.3, though, so far, no theoretical justification can be found for this strategy. This chapter aims at investigating and validating this two-step bootstrap approach.

The ARAB extends and incorporates both the ordinary moving block bootstrap and the residual bootstrap. These two bootstrap approaches each seem to be the most extreme ones in contrary directions. While the MBB works under very mild assumptions and is nearly applicable to any weakly stationary sequence, its performance is very sensitive to the choice of the block length l. On the other hand, the residual bootstrap aims at pinpointing an underlying parametric model, consider e.g. an autoregressive (AR) model, and then resampling the i.i.d. innovations. This approach delivers remarkably good performances but is crucially restricted to the process class considered. Our intention in this study is to incorporate the advantages of these different proposals and to weaken their niggles into one new procedure. We aim at an as widely applicable as possible procedure that performs as robust and satisfying as possible.

The remainder of the chapter is as follows. Section 5.2 introduces some technical preliminaries and then describes the ARAB procedure. Its connection to already established bootstrap proposals will be further discussed. Section 5.3 states the main theorem of this chapter that the ARAB is valid under some mild assumptions for the large class of functions of generalized means. Furthermore, discussions on the second-order correctness will conclude the section. Section 5.4 discusses the choice of the free parameters, namely the order of the model fit p and the length of the blocks l. The results from several simulation studies are reported in Section 5.5. Section 5.6 shortly concludes. Technical details and proofs are referred to the proof section.

5.2 AR-aided block bootstrap

5.2.1 Preliminaries

Throughout this chapter, the following basic assumptions are imposed on the underlying stochastic process $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$.

Assumption 4. *Let $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$ be a zero mean (strictly) stationary process with $E|X_t|^{2+\delta} < \infty$ for some $\delta > 0$. The autocovariance function $\gamma_X : \mathbb{Z} \to \mathbb{R}$, where $\gamma_X(h) = EX_tX_{t+h}$ is assumed to fulfill $\gamma_X(0) > 0$ as well as $\sum_{h\in\mathbb{Z}} |\gamma_X(h)| < \infty$.*

This assumption ensures that a continuous spectral density f_X of the underlying process $\mathbf{X}$ exists

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_X(h) e^{-ih\omega}, \quad \omega \in (-\pi, \pi]. \tag{5.1}$$

The assumed absolute summability of the autocovariance function excludes processes with unbounded spectral density, as for example long memory time series and as a special case fractionally integrated autoregressive moving average time series.

Sometimes we indicate by an index to which time series the autocovariance function or spectral density belongs, i.e. we write γ_X and f_X.

In this chapter we focus on statistics which are functions of generalized means only. In the context of bootstrap approaches, this class of statistics is often considered since Künsch (1989) (see his Example 2.2.). It includes e.g. versions of the sample mean, sample autocorrelations, sample autocovariances and the Yule-Walker estimates in autoregressive (AR) models. Having observations $X_1, \ldots, X_n$ at hand the estimator of interest T_n for the parameter $f(\theta)$ then is given by

$$T_n = f\left(\frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t, \ldots, X_{t+m-1})\right), \tag{5.2}$$

where $g = (g_1, \ldots, g_q)'$ and $g : \mathbb{R}^m \to \mathbb{R}^q$ for some $m \in \{1, \ldots, n\}$ and $f : \mathbb{R}^q \to \mathbb{R}^{\tilde{q}}$, where $q, \tilde{q} \geq 1$. In the following, we want to investigate bootstrap possibilities for and to mimic the limiting behavior of (5.2). For that, we assume a central limit theorem to hold.

Assumption 5. *It holds*

$$\frac{1}{\sqrt{n-m+1}} \sum_{t=1}^{n-m+1} \left(g(X_t, \ldots, X_{t+m-1}) - \theta\right) \xrightarrow{\mathcal{D}}_{n\to\infty} \mathcal{N}\left(0_q, \Sigma_{q\times q}\right), \tag{5.3}$$

where $(\Sigma_{q\times q})_{u,v} = \sum_{k=-\infty}^{\infty} Cov(g_u(X_0, \ldots, X_{m-1}), g_v(X_k, \ldots, X_{k+m-1}))$*, for* $u, v = 1, \ldots, q$*, which implies that*

$$\sqrt{n}\left(T_n - f(\theta)\right) \xrightarrow{\mathcal{D}}_{n\to\infty} f'(\theta)\, \mathcal{N}\left(0_q, \Sigma_{q\times q}\right), \tag{5.4}$$

where $f'(\cdot) := Df(\cdot)$ *is the Jacobi matrix of the function* f*.*

For such limiting results to hold one usually requires f and g to satisfy some smoothness conditions and the stochastic process to fulfill some weak dependence properties, e.g. such as mixing conditions. While the conditons on the functions f and g determine the limiting covariances, only the dependence structure – namely the autocovariance – of the underlying series $\mathbf{X}$ influences our results. The specific weak dependence properties have no effect. Carlstein (1986) discusses this characteristic for α-mixing series where the mixing coefficients do not affect the limiting results. That is the reason why we do not constrain our investigations to mixing processes and keep the assumptions as weak and as general as possible. Nevertheless, we require some conditions on the functions (compare Bühlmann (1997)):

Assumption 6. *The function* $f = (f_1, \ldots, f_{\tilde{q}})$ *has continuous partial derivatives* $y \mapsto \sum_{i=1}^{q} \frac{\partial f_u}{\partial x_i}\Big|_{x=y} y_i$*, for all* $u = 1, \ldots, \tilde{q}$*,* $i = 1, \ldots, m$*, for* y *in a neighborhood of* θ *and the differentials at* θ*,* $y \mapsto \sum_{i=1}^{q} \frac{\partial f_u}{\partial x_i}\Big|_{x=\theta} y_i$*, for all* $u = 1, \ldots, \tilde{q}$*, do not vanish. Further, the function* g *has continuous partial derivatives of order* h*,* $h \geq 1$*,* $\frac{\partial^h g_i(x)}{\partial x_{i_1} \ldots \partial x_{i_h}}\Big|_{x=y}$ *which satisfy the following Lipschitz condition:*

$$\left| \frac{\partial^h g_i(x)}{\partial x_{i_1} \ldots \partial x_{i_h}}\Big|_{x=y} - \frac{\partial^h g_i(x)}{\partial x_{i_1} \ldots \partial x_{i_h}}\Big|_{x=z} \right| \leq C_u \|y - z\|, \tag{5.5}$$

where $u = 1, \ldots, q$*;* $1 \leq i_1, \ldots, i_h \leq m$*,* C_u *is some suitable constant and* $\|\cdot\|$ *denotes the Euclidean norm, for every* $x, y, z \in \mathbb{R}^m$*.*

Assumptions 5 and 6 are in line with the assumptions in Götze and Künsch (1996) and Bühlmann (1997). Furthermore, it should be noted that, here, we do not require a specific type of dependence like e.g. an autoregressive model. Especially, no further assumptions on e.g. suitable mixing conditions are made.

5.2.2 The AR-aided block bootstrap algorithm

The autoregressive-aided block bootstrap procedure proposes to generate bootstrap replicates T_n^* of the estimator T_n which rely on bootstrap pseudo-time series $X_1^*, \ldots, X_n^*$. For the remainder, the block length may be denoted by l and let the total number of blocks b be defined as the smallest integer with $bl \geq n + p$. The bootstrap procedure consists of the following steps:

Step 1:
Given the observations $X_1, \ldots, X_n$ we fit an autoregressive process of order p, where p may depend on the particular sample at hand. The AR order p could be random or chosen according to an order selection criterion like Akaike's information criterion (AIC) or final prediction error (FPE), see Section 5.4. This leads to estimated autoregressive parameters $\hat{a}_1(p), \ldots, \hat{a}_p(p)$ which are obtained from the common Yule-Walker estimates; compare Brockwell and Davis (1991):

$$(\hat{a}_1(p), \ldots, \hat{a}_p(p))' = \underset{(c_1,\ldots,c_p)}{\operatorname{argmin}} \sum_{t=p+1}^{n} \left(X_t - \sum_{j=1}^{p} c_j X_{t-j} \right)^2 . \tag{5.6}$$

Consider the resulting pseudo-residuals

$$\hat{U}_t = X_t - \sum_{j=1}^{p} \hat{a}_j(p) X_{t-j} \quad \forall t = p+1, \ldots, n. \tag{5.7}$$

Step 2:
Center the estimated residuals $\hat{U}_{p+1}, \ldots, \hat{U}_n$ at $\bar{\hat{U}}$ which is defined by

$$\bar{\hat{U}} := \frac{1}{n-p-l+1} \left[\sum_{i=1}^{n-p} \hat{U}_i - \sum_{i=1}^{l-1} \frac{i}{l} \hat{U}_{l-i} - \sum_{i=1}^{l-1} \frac{i}{l} \hat{U}_{n-p-l+1+i} \right]. \tag{5.8}$$

Note that this is not the usual mean but assures mean zero according to the moving block bootstrap distribution which is applied in the next step. We refer to Bickel and Freedman (1981) or Davison and Hinkley (1997) for discussions on exchanging the usual mean by (5.8). Denote $\hat{U}_t^c = \hat{U}_t - \bar{\hat{U}}$, for all $t = p+1, \ldots, n$.

Step 3:
Generate bootstrap residuals $\left(U_{1-p}^*, U_{2-p}^*, \ldots, U_n^*\right)$ by application of the moving block bootstrap to the sequence $\hat{U}_{p+1}^c, \ldots, \hat{U}_n^c$. Thus, draw with replacement b i.i.d. random

variables $i_1, \ldots, i_b$ *having discrete uniform distribution on the set* $\{p, \ldots, n-l\}$ *and obtain bootstrap pseudo-residuals*

$$\left(U_{1-p}^*, \ldots, U_n^*\right) = \left(\hat{U}_{i_1+1}^c, \ldots, \hat{U}_{i_1+l}^c, \hat{U}_{i_2+1}^c, \ldots, \hat{U}_{i_1+1}^c, \ldots, \ldots, \hat{U}_{i_b+l}^c\right). \tag{5.9}$$

For technical reasons throughout the proofs we further define $U_t^* \equiv 0$ *for all* $t < 1-p$.

Step 4:
Compute the bootstrap time series via

$$X_t^* = \sum_{j=1}^{p} \hat{a}_j(p) X_{t-j}^* + U_t^* \quad \forall t = 1, \ldots, n, \tag{5.10}$$

and finally obtain the bootstrap estimate:

$$T_n^* = f\left(\frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t^*, \ldots, X_{t+m-1}^*)\right). \tag{5.11}$$

Some remarks are now in order. It is important to note that we do not attempt to reduce the time series to i.i.d. values and so the present approach distinguishes from purely residual based techniques (e.g. Davis (1977), Freedman (1984), Efron and Tibshirani (1986), Bose (1988)). These techniques could only be used if and only if the underlying time series $\mathbf{X}$ is an AR process of order p with i.i.d. noise sequence. If the true order is misestimated then the usual residual-based approaches may fail. In here, the main goal of fitting an AR model is in the spirit of the so-called *prewhitening* idea (Press and Tukey (1956)). As Kreiss and Paparoditis (2003) pointed out, AR fits are able to catch dominant peaks of the underlying spectral density remarkably well. The remainder part of the time series, that is the process $\left\{\hat{U}_t : t \in \mathbb{Z}\right\}$, then is somehow *whitened* and its properties are (hopefully) easier to address by the block bootstrap. Even in the case that the underlying process is an AR process and we misestimated its order, the present proposal remains valid, because the block bootstrap can address for non-detected structures of the AR fit. (An exemplary visualization of the prewhitening effect is given by Figure 5.3 in the numerical section.)

Understanding such a prewhitening as a graphical device, the AR order p is selected over a suitable set of values for which the smoothed rescaled periodogram is closest to constant.

It is worth mentioning that fitting an autoregression should be seen as a (convenient) example. Of course fitting other parametric models may also be considered. Furthermore, additional steps of fitting diverse parametric models one after another could further improve performances but are not considered in this study.

The use of Yule-Walker estimators is rather convenient since it guarantees – among further nice abilities – that the complex polynomial $A(z) = 1 - \sum_{j=1}^{p} \hat{a}_j(p) z^j$ has no roots on or within the unit circle $\{c \in \mathbb{C} : |c| \leq 1\}$, i.e. the bootstrap process always constitutes a stationary process. However, it is not necessary to work with Yule-Walker estimates; any $\sqrt{n}$-consistent parameter estimate would suffice.

The centering step is with respect to the ensuing moving block bootstrap. The usual mean does not occur as the expected value because of the overlapping of the blocks. If non-overlapping blocks were used, one would have to modify the centering step. However, Lahiri (1999) found that using overlapping blocks is to be preferred over non-overlapping blocks and that using random block lengths commonly leads to mean-squared errors larger than those for non-random block lengths. The accuracy of the block bootstrap critically depends on the block size that must be supplied by the user. For discussions on the block size see Section 5.4. Additional references for various aspects of block bootstrap methods are e.g. Andrews (2004), Bühlmann (1994), Carlstein et al. (1998), Davison and Hall (1993), Dowla et al. (2003), Hall et al. (1995) and Paparoditis and Politis (2002).

5.3 Bootstrap validity

In this section we state our main result which will show that the ARAB is valid for functions of generalized means.

Recall that the ARAB allows for the generation of pseudo-time series $X_1^*, \ldots, X_n^*$ via (5.10). The estimated (possibly complex-valued) autoregressive polynomial used, namely $1 - \sum_{j=1}^{p} \hat{a}_j(p) z^j$, has a counterpart for the original underlying time series $\mathbf{X}$. Denote this polynomial by

$$A(z) = 1 - \sum_{j=1}^{p} a_j(p) z^j, \tag{5.12}$$

where the coefficients $a_j(p)$, $j = 1, \ldots, p$, may denote the most-likely autoregressive

parameters with respect to the Yule-Walker proposal which in this case coincides with the best least squares approximation. The polynomial $A(z)$ has no complex roots on or within the unit circle due to the use of the Yule-Walker estimates. For computational reasons it is much more likely to use the corresponding inverse polynomial

$$A^{-1}(z) = \left(1 - \sum_{j=1}^{p} a_j(p) z^j\right)^{-1} = \sum_{j=0}^{\infty} \psi_j(p) z^j =: \Psi(z), \tag{5.13}$$

for $z \in \mathbb{C}$, and where always $\psi_0(p) = 1$, and the same holds for its estimated counterpart. In the following we will equivalently use the notation ψ_j for $\psi_j(p)$ and $\hat{\psi}_j$ for $\hat{\psi}_j(p)$.

The inverse autoregressive polynomial $\Psi(z)$ as defined in (5.13) is bounded away from zero for all $|z| \leq 1$, $z \in \mathbb{C}$, and the coefficients fulfill $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$ for any $r \geq 0$. This directly follows by Cauchy's inequality for holomorphic functions (e.g. Fischer and Lieb (1988), Theorem 6.1). Thus, models with polynomial decay of the moving average coefficients are included. Autoregressive moving-average (ARMA) models of finite order satisfy this characteristic with exponential decay of the coefficients. It is worth emphasizing again that no underlying AR model is supposed to exist and thus, the pseudo-time series $\{U_t : t \in \mathbb{Z}\}$ is not assumed to fulfill neither white noise nor i.i.d. conditions. However, we require the following mild assumption on the time series $\{U_t : t \in \mathbb{Z}\}$ and its corresponding estimates.

Assumption 7. *The remainder process $\{U_t : t \in \mathbb{Z}\}$, obtained by application of an autoregressive fit of order p to the time series $\{X_t : t \in \mathbb{Z}\}$, fulfills*

$$E\left[U_t^{\alpha}\right] < \infty \quad \forall \alpha \leq 2(h+1), \tag{5.14}$$

and the empirical moments converge towards their theoretical counterparts

$$\frac{1}{n-p} \sum_{t=1}^{n-p} \hat{U}_t^{\alpha} \xrightarrow{P} E\left[U_t^{\alpha}\right] \quad \forall \alpha \leq 2(h+1). \tag{5.15}$$

(See Assumption 6 for a definition of h.)

In this thesis, our investigations are restricted to weak dependent (or mixing) stochastic processes which allow for a further mild assumption.

Assumption 8. *For the autocovariance function it holds*

$$\sqrt{n}(\widehat{\gamma}_X(h) - \gamma_X(h)) = \mathcal{O}_P(1) \ \forall h = 0, \ldots, p, \tag{5.16}$$

where $\widehat{\gamma}(h)$ denotes the empirical autocovariance to the lag h.

The following theorem states that the autoregressive-aided block bootstrap is asymptotically consistent for the distribution of functions of generalized means. In particular, because the ARAB imitates accurately the weak dependence of the underlying process, it manages to reproduce correctly the asymptotic variance for this class of statistics.

Theorem 5.1. *Suppose that Assumptions 4–8 hold with $\delta = 2h$. (See Assumption 6 for a definition of h.) T_n^* may be defined as in (5.11), and $\Sigma_{q\times q}$ as in (5.3). Further assume that $l \to \infty$ and $\frac{l^{2+2/\delta}}{n} \to 0$, as $n \to \infty$. Then, denoting $\theta^* = E^* g(X_t^*, \ldots, X_{t+m-1}^*)$, and as $n \to \infty$, it holds...*

(i) ...the limiting covariance matrix:

$$(n-m+1)^{-1} \, Var^* \left(\sum_{t=1}^{n-m+1} \left(g(X_t^*, \ldots, X_{t+m-1}^*) - \theta^* \right) \right) \to \Sigma_{q\times q}. \tag{5.17}$$

(ii) ...the central limit result:

$$\sqrt{n}(T_n^* - f(\theta^*)) \xrightarrow{\mathcal{D}} f'(\theta) \, \mathcal{N}\left(0_q, \Sigma_{q\times q}\right). \tag{5.18}$$

(iii) ...bootstrap validity:

$$\sup_{x\in\mathbb{R}^{\tilde{q}}} \left| P^*(\sqrt{n}(T_n^* - f(\theta^*)) \leq x) - P(\sqrt{n}(T_n - f(\theta)) \leq x) \right| = o_P(1). \tag{5.19}$$

This theorem continues a series of well-established results. Earlier, Theorem 3.5 of Künsch (1989) extended the results of Singh (1981), Theorem 1.A., and Bickel and Freedman (1981), Theorem 2.1. The above presented Theorem 5.1 further generalizes Theorem 3.5 of Künsch (1989) by allowing for model fits as a pre-step.

Remark 5.2. *Bootstrap validity holds as long as $\frac{l^{2+2/\delta}}{n} \to 0$ for $n \to \infty$ is fulfilled. The parameter δ is thereby depending on the time series $\mathbf{X}$ and is the same value as for $E|X_t|^{2+\delta} < \infty$ in Assumption 4. Then the condition in Theorem 5.1 holds if $l = n^{1/2-\varepsilon}$ for some ε yielding $\frac{1}{2(1+\delta)} < \varepsilon < \frac{1}{2}$. This means that the higher moments of the original time series exist the shorter the rate of the blocks may be chosen.*

Remark 5.3. *It should explicitly emphasized that the ARAB is tailor-made but not limited to AR processes with non-independent innovation sequences, so called weak AR or weak ARMA processes. Such processes arise quite naturally and in several situations: Strong ARMA processes with non-normal distributed noise observed at lower frequencies yield these properties (cf. Niebuhr and Kreiss (2014), or see Chapter 2). Another example is given when observing only the first component of a vector-valued AR process. Furthermore, equidistant samples of continuous-time ARMA processes also fulfill some weak AR representation (Brockwell et al. (2010), or Chapter 3). However, the ARAB is wider applicable than to only these classes of processes. Due to the block bootstrap mechanism the procedure is as widely applicable as the MBB is.*

Remark 5.4. *In general, the bootstrap expected value θ^*, or $f(\theta^*)$ respectively, may differ from the sample-based statistic. Though there are given many examples in the literature where these values coincide, e.g. for i.i.d. data and Efron's bootstrap in case of the sample mean, this assertion usually does not hold for block-based approaches (cf. Bickel and Freedman (1981), Davison and Hinkley (1997), along with others). Even for the simple case of the sample mean the MBB's expected value computes to*

$$E^*\left[X_t^*\right] = \frac{1}{n-l+1}\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{l-1}\frac{i}{l}X_{l-i} - \sum_{i=1}^{l-1}\frac{i}{l}X_{n-l+1+i}\right], \tag{5.20}$$

which obviously does not coincide with the simple sample mean $\frac{1}{n}\sum_{t=1}^{n} X_t$ for block lengths larger than 1. The computation of the bootstrap expected value is usually not intuitive. Though while theoretical results are derived using the true bootstrap expected values, in application cases estimation biases may occur.

The present investigations consider the case when the observations are short-range dependent, i.e. they satisfy some form of mixing conditions (strong mixing, uniform mixing) with a rapidly decaying mixing coefficient. Under long-range dependence the situation may be different and the MBB and thus the ARAB may fail (Lahiri (1993)). Several asymptotic refinements of the block bootstrap under short-range dependence are considered by Lahiri (1992), Götze and Künsch (1996), and Lahiri (1999), while Kim and Nordman (2011) among others investigated the block bootstrap under long-range dependence. For further references on the validity of the block bootstrap see Lahiri (2003) and the references therein.

5.3.1 On second-order properties

The first order correctness for the MBB follows essentially from the convergence of the bootstrap sample covariance matrix to the limiting covariance matrix (Künsch (1989)). Following Götze and Künsch (1996) it is essential to center the bootstrap estimate by its true parameter θ^* – which does not coincide with the estimator relying on the time series T_n – to obtain second-order correctness. The MBB then provides second-order correct approximations to statistics that are smooth functions of sample means (Lahiri (1991), Götze and Künsch (1996)). The essential reason for second-order correctness is the asymptotically correct skewness of the blockwise bootstrap distribution (Künsch (1989)). We should note that the class of smooth functions of sample means is included in the class of functions of generalized means (5.2) considered in this chapter. Since the Yule-Walker estimates result in well-fitted models and the fits (at least) *prewhiten* the time series (Tukey (1958)), we conjecture that the ARAB would yield second-order correctness at least under the framework of a smooth function model, although a rigorous proof is well beyond the scope of this thesis.

5.4 Choice of the parameters

Before proceeding with numerical examples of the finite-sample performance of our bootstrap procedure, some remarks on the choice of the parameters p and l are given. We restrict our discussion to some rather heuristic ideas on how to choose these parameters. These may be helpful guidelines in applications. We do not claim any asymptotic optimality for either approach. The aim is rather to provide sensible answers for small to moderate sample sizes. Clearly, more theoretical results are required in order to make definite recommendations.

5.4.1 Choice of the autoregressive order

The autoregressive order p may be chosen data dependent, that is, p could be either randomly chosen or according to an order selection criterion. The most well-known solutions for the order selection problem include the final prediction error (FPE) (Akaike (1969)), Akaike information criterion (AIC) (Akaike (1973), (1974)), and minimum description length (MDL). For a deeper discussion see Liang et al. (1993).

Of course, higher model orders provide better fits but it is important to avoid an *overfitting* to the present data set. The motivation is to balance the complexity of the model and the volatility of the resulting residuals. The possibly most convenient way to address this conflict is according to the AIC, which minimizes the function

$$AIC(p) = \underset{p}{\operatorname{argmin}} \left\{ \hat{\sigma}^2(p) \left(1 + \frac{2p}{n} \right) \right\} \tag{5.21}$$

over a range of $p = 1, 2, \ldots, p_{max}(n)$, where $\hat{\sigma}^2(p)$ describes the model variance. The maximum value for the order $p_{max}(n)$ may be defined by e.g. $p_{max}(n) = 10 \log_{10}(n)$, which is the default value in S-PLUS. For processes obeying an infinite-order AR structure, Shibata (1981) found that such choice leads to an asymptotically optimal AR spectral density estimator. This is no more valid if the underlying process does not obey this structure. In this case, an appropriate AR fit can rely on the idea of prewhitening.

Broersen and de Waele (2004) stated that AIC tends to select orders that are too high, even asymptotically, and even earlier Hannan (1980) slightly modified the AIC to assure consistency. However, for finite sample sizes we suppose the use of AIC.

Remark 5.5. *It should be emphasized that usual residual and model-based bootstrap approaches assume that the user has correctly specified the order of the underlying autoregessive model. In our setup this crucially is different. We neither assume an autoregressive representation of the process* $\mathbf{X}$ *nor - in the case that* $\mathbf{X}$ *has an AR representation - that the correct order is identified. The model fit just prewhitens the data and prepares for the ensuing block bootstrap. Even in the case that* $\mathbf{X}$ *has an AR representation it is not essential to identify the correct order. Misspecification of the order will not alter the approach's validity. For that reason the ARAB can be seen as a robustification of the ordinary residual based bootstrap.*

5.4.2 Choice of the block length

The main practical problem in applying the block bootstrap lies in choosing the block length l. This problem is shared by all blocking methods, such as, for example, subsampling approaches (e.g. Politis et al. (1999)). The asymptotic conditions, at least for first order theory, are usually $l \to \infty$ and $l/n \to 0$ as $n \to \infty$. Although any choice of l satisfying these conditions will yield the required consistency, the two

asymptotic conditions do not give much guidance with respect on how to choose l when faced with a finite sample. Indeed, for l too close to n all bootstrap statistics will be almost equal to the estimator, resulting in the bootstrap distribution being too tight and in undercoverage of confidence intervals. Lahiri (1999) makes this intuition precise by proving, in the context of smooth functions of means, that for $l/n \to 1$, the approximation collapses to a point mass at zero. On the other hand, if l is too small, the intervals can undercover or overcover depending on the state of nature. Lahiri deeply considered the optimal block length for the class of smooth functions of means in several studies (e.g. Lahiri (2003)).

The class of smooth functions of means is included in the class of functions of generalized means (5.2), as one would expect. One obtains this subclass by setting $m = 1$ and $\tilde{q} = 1$ and allowing for vector-valued time series. Optimal block lengths are often considered only for this subclass. Lahiri et al. (2007) and further publications focus on minimizing the mean squared error of the statistic of interest and result in optimal rates of $l = const. \cdot n^{-1/3}$. The constant belonging to $n^{-1/3}$ is a function of the considered statistic and the dependence among the observations. Those constants are typically hard to pin-point and often unknown to the user. We recall the optimal block length l^{opt} for smooth functions of means as found by Lahiri (2003), Corollary 7.1, for the bias functional (the variance functional is of very similar form) since it will bring insight for further discussion.

Corollary 5.6 (Lahiri (2003)). *Let Z_∞ be a q-dimensional Gaussian random vector with mean zero and covariance matrix $\Sigma_\infty = \sum_{j=-\infty}^{\infty} |j| E(X_1 - \mu)(X_{1+j} - \mu)$, and $c_\alpha = D^\alpha f(\theta)/\alpha!$, $D^\alpha = \frac{\partial^{\alpha_1 + \cdots + \alpha_q}}{\partial x_1^{\alpha_1} \cdots \partial x_q^{\alpha_q}}$, $\alpha! = \prod_{i=1}^{q} \alpha_i!$ for $\alpha = (\alpha_1, \ldots, \alpha_q)' \in \mathbb{Z}_+$; then*

$$l^{opt}(n) = n^{1/3}(2A^2/v^2)^{1/3} + o(n^{1/3}), \tag{5.22}$$

where

$$A = -\sum_{|\alpha|=1} \sum_{|\beta|=1} c_{\alpha+\beta} \left[\sum_{j=-\infty}^{\infty} |j| E(X_1 - \mu)^\alpha (X_{1+j} - \mu)^\beta \right] \tag{5.23}$$

and

$$v^2 = \frac{2}{3} Var \left(\sum_{|\alpha|=2} c_\alpha Z_\infty^\alpha \right). \tag{5.24}$$

The optimal rate of $l = const. \cdot n^{-1/3}$ may change when considering different estimates, such as e.g. the distribution function or quantiles (Hall et al. (1995), Lahiri (1999), Politis et al. (1999)). However, for this discussion we focus on smooth functions of means. Nordman and Lahiri (2014) recently investigated the convergence rates of empirical block length selectors under the framework of the smooth function model (cf. Hall (1992)). They compared general nonparametric methods for estimating optimal block lengths, namely the ones of Hall et al. (1995) and of Lahiri et al. (2007). In general, the true optimal block size and the optimal block size determined by these approaches differ by a margin of $\mathcal{O}\left(n^{-1/3}\right)$ on the relative scale, which should be seen as an upper bound on their accuracy. Under certain dependence (mixing) conditions these rates can be improved to $\mathcal{O}\left(n^{-2/7}\right)$ for the Lahiri et al. (2007) approach and to $\mathcal{O}\left(n^{-1/6}\right)$ for the Hall et al. (1995) method. However, since we do not necessarily assume for mixing sequences this work contents itself with considering the more general case.

Remark 5.7. *Regarding (5.22), it is obvious that the optimal rate of $n^{1/3}$ will not change if we investigate the prewhitened pseudo-residuals $\left\{\hat{U}_t : t \in \mathbb{Z}\right\}$ instead of the original time series $\mathbf{X}$. Thus, the optimal rate for the autoregressive-aided block bootstrap results to $l = const. \cdot n^{1/3}$. Nevertheless, the occuring constant will change of course. Having (5.23) at hand, it becomes clear that if X_t is replaced by the prewhitened - and thus closer to white noise - process $\hat{U}_t$, the constant A will decrease depending on the goodnis of the autoregressive fit. Intuitively, this is exactly what one would expect. Roughly speaking, the prewhitening step is able to take out some dependence of the series, and consequently, the pseudo-residuals $\hat{U}_t$ will be less dependent than the original series $\mathbf{X}$. Because of this one might naively think that shorter blocks would be sufficient. As a guideline, we suppose that ordinary block length choices remain suitable and can be applied to $\hat{U}_t$. Further approaches on choosing optimal block lengths are stated by several authors using different optimality criteria. We refer the reader to Hall et al. (1995), Bühlmann and Künsch (1999), Politis and White (2004), Lahiri et al. (2007) or Patton et al. (2009).*

5.5 Numerical examples

In previous sections, we investigated the ARAB from a theoretical perspective. To sustain our theoretical findings, here we investigate via simulations the finite-sample performance of the autoregressive-aided block bootstrap (ARAB) in comparison

with the residual based bootstrap, from now called AR, and the ordinary moving block bootstrap (MBB). The MBB and the ARAB are applicable to any stationary (weak dependent) time series while the AR bootstrap may fail when leaving the class of pure autoregressions with i.i.d. innovations. Anyway, for this class the AR bootstrap is tailor-made.

If it is not indicated otherwise, we will consider the following simulation pattern. As a quantity of interest of the distribution

$$\sqrt{n}(\hat{\gamma}_X(2) - \gamma_X(2)) \tag{5.25}$$

we choose several quantiles, namely the 1%, 5%, 50% (median), 90%, 95% and the 99% percentile. From the bootstrap we present boxplots of the obtained bootstrap quantiles in comparison with the true finite sample quantiles. Within the figures, the true finite sample values are usually highlighted in red color.

The present studies are based on $K = 500$ independently simulated (Monte-Carlo) time series samples. For each sample $B = 800$ independent bootstrap repetitions were simulated from which the bootstrap estimates were computed. The sample size was set to $n = 200$ and for both the MBB and the ARAB several different block lengths $l = 6, 12, 24, 36$ were chosen. It is worth noting that for the specified sample size $n = 200$ it holds $\left\lceil n^{1/3} \right\rceil = 6$. The true finite sample values are gained by simulation of 20.000 cases. We distinguish several scenarios with different intentions at hand.

Remark 5.8. *As previously discussed in Remark 5.4 the bootstrap expectation may differ from the sample-based estimator which may lead to biases. Explicit computation of these biases is quite challenging and rarely investigated. The question for bias correction thus is present but hard to pin-point. We will not give any detailed computational results on this topic but want to emphasize on this rather relevant problem. It is somehow charming to use the ordinary sample-based estimate and especially for longer block sizes the bias should become small. Indeed, this is what the ensuing figures show. Nevertheless, exceptions might arise in some cases where even for smaller block sizes less biased results may be found. We follow the ordinary application and will approximate (5.25) by*

$$\sqrt{n}(\hat{\gamma}^*_X(2) - \hat{\gamma}_X(2)) \tag{5.26}$$

instead of using the true bootstrap expectation $\gamma_X^*(2)$*. Of course, bias effects will arise but become small for suitable block lengths.*

5.5.1 AR time series

First, we consider autoregressive time series with i.i.d. innovations. For this class of processes the AR bootstrap is tailor-made, and thus one might expect the AR to outperform other approaches. The model fit in the ARAB should result in improvements in relation to the MBB. For that reason, intuitively, we expect the ARAB to perform in between the MBB and the AR. The simulation study was done for the AR(2) process

$$X_t = -0.9X_{t-1} - 0.7X_{t-2} + e_t, \tag{5.27}$$

where e_t was chosen as standard normal distributed. The simulation results are summarized in Figures 5.1 and 5.2.

Under usage of AIC order selection, one is lead to the true order $p = 2$ in 375 of 500 cases. Figure 5.1 compares boxplots of the bootstrapped quantiles to the true distribution quantiles (dashed red lines). It can easily be seen that the AR and the ARAB bootstrap capture the goal values remarkably well while the MBB shows stronger sensitivity to the choice of the block length. Anyhow, when the block length is appropriately chosen the MBB performs satisfying. The dependence might be too strong for the MBB to be captured by short blocks.
Another observable aspect is the increasing variability for increasing block lengths. Compare the AR – which indeed is an ARAB bootstrap with block length $l = 1$ – to the ARAB boxplots and regard the movement of the 95% whiskers covering wider and wider ranges as the block length increases.

Figure 5.2 visualizes the finite sample distribution with bootstrapped 95% confidence bands for the single methods. These plots should be read as horizontal pointwise confidence intervals to the confidence level 95%. In the first row the AR-aided approaches are shown, all performing very well and yielding the confidence ranges. The second row gives the corresponding results for the MBB which also are satisfying if the block size is chosen large enough. Comparing the MBB with the ARAB, the ARAB is much less sensitive to the block length than the MBB. Clearly, this is due to the prewhitening model-fit.

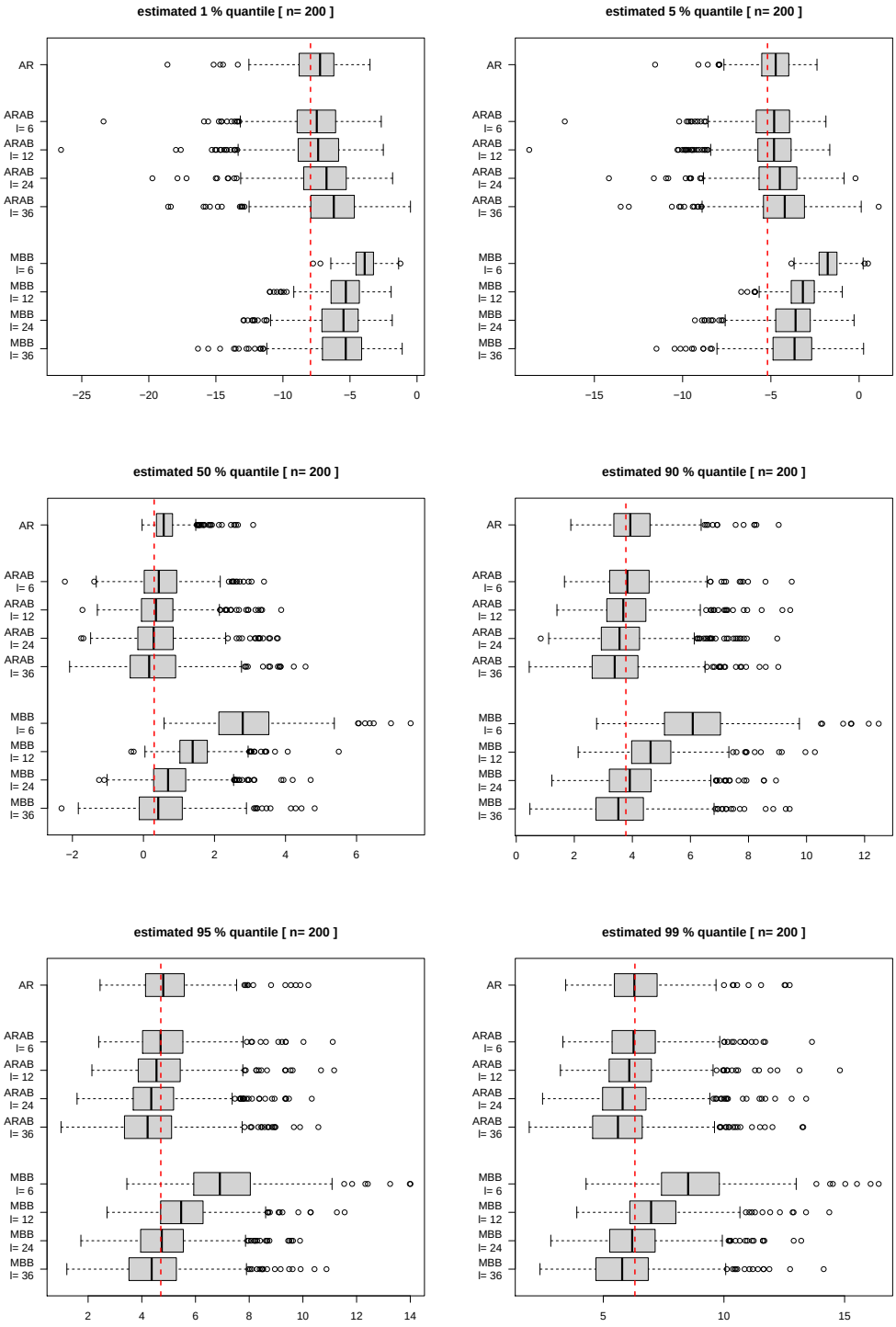

Figure 5.1: AR time series (using AIC): boxplots of bootstrapped quantiles of the AR bootstrap, the ARAB bootstrap and MBB bootstrap with true finite sample quantile (dashed red line).

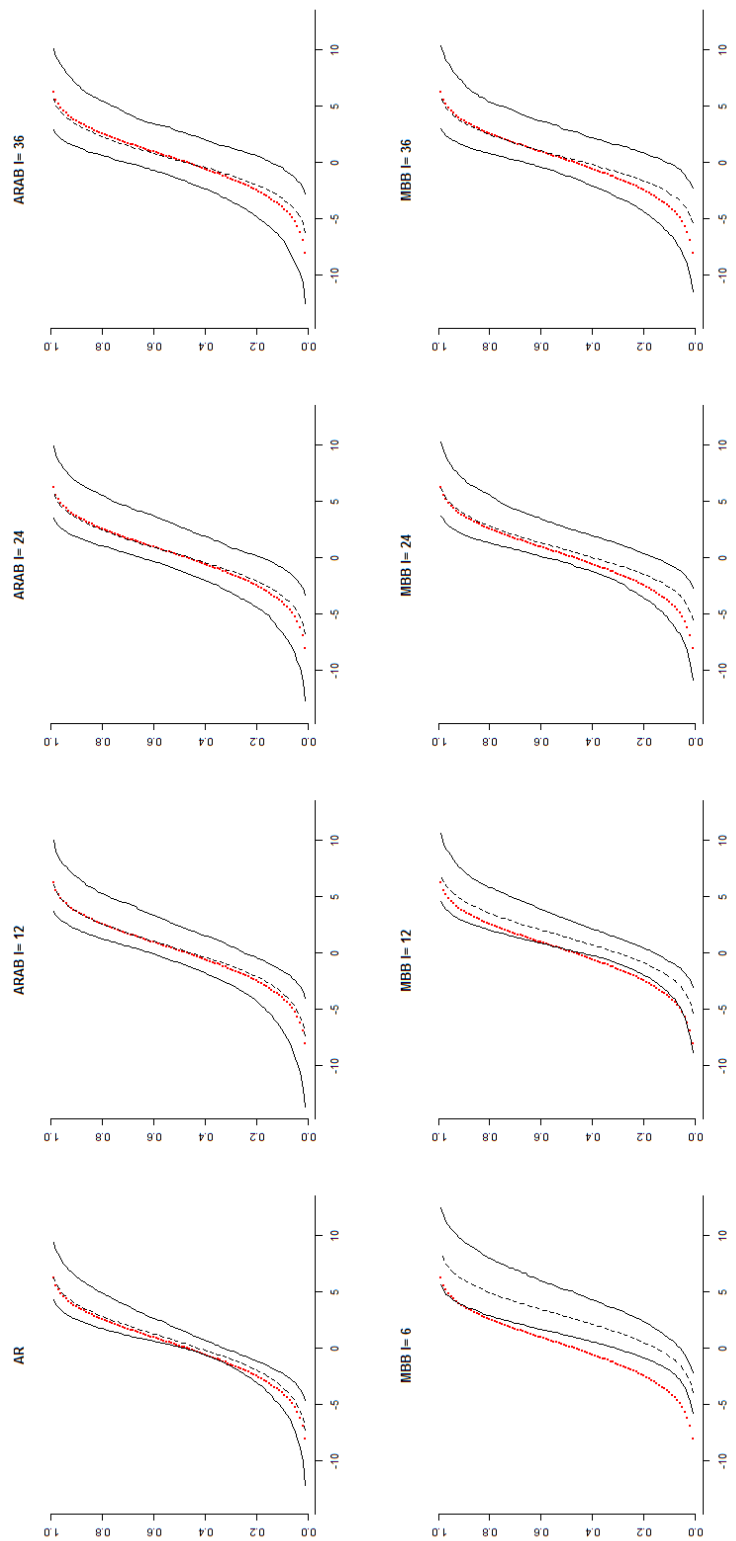

Figure 5.2: AR time series (using AIC): Finite sample distribution (dotted red line) in comparison to 95% confidence bands with dashed median of the AR bootstrap (first row, first plot), the ARAB bootstrap (first row, second to fourth plots) and the MBB (second row).

5.5.2 AR Time Series with misspecified order

It might be of interest to regard the consequences of misspecifying the underlying model order. A simulation study was done for the AR(2) process as in equation (5.27) where a fixed model order of 1 was suggested instead of using AIC. If the model fit order is not chosen by AIC but by a priori fixation, the prewhitening effect may be very small and may not result in visual improvements. Furthermore, the AR bootstrap might substantially go wrong. Figure 5.4 underlines this suggestion. It seems that the ARAB and the MBB perform comparable. The arising question is then why there is no improvement due to the model fit. Having a look at the spectral densities of the original AR series $\mathbf{X}$ and the pseudo residuals $\hat{U}_t$ relying on the fixed model fit orders gives insight (see Figure 5.3). A too small model fit may – depending on the model coefficients – not be able to whiten the original time series essentially. However, while the effect may not be very large the spectral density becomes closer to constant. Thus, the performance of the ARAB should at least not be worse than the performance of the MBB.

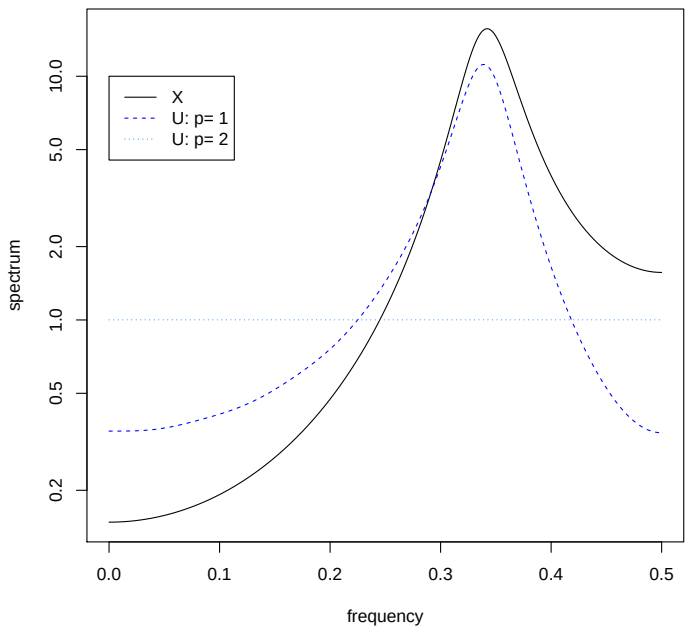

Figure 5.3: AR(2) spectral density compared to spectral densities of pseudo residuals for model fit orders 1 and 2.

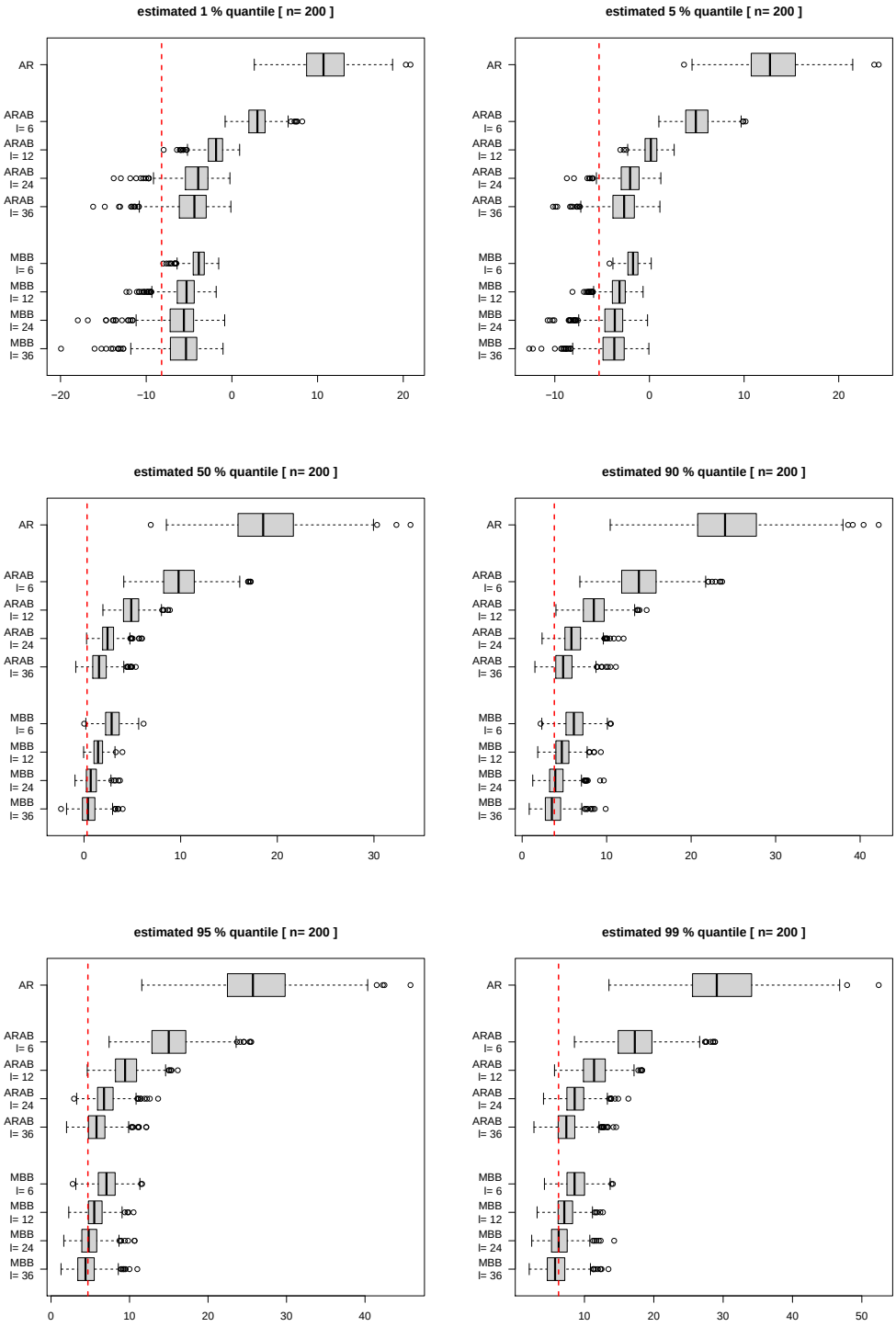

Figure 5.4: AR time series (fixed model 1 for model fit): boxplots of bootstrapped quantiles of the AR bootstrap, the ARAB bootstrap and MBB bootstrap with true finite sample quantile (dashed red line).

5.5.3 ARMA Time Series

Another simulation study was done for the following ARMA(4,3) process

$$\begin{aligned} X_t &= -0.4X_{t-1} + 0.3X_{t-2} + 0.2X_{t-3} - 0.1X_{t-4} \\ &\quad + e_t + 0.8e_{t-1} + 0.7e_{t-2} + 0.9e_{t-3}, \end{aligned} \tag{5.28}$$

where e_t was chosen as standard normal distributed. In this simulation scenario we know from theory that the AR bootstrap should fail and this is exactly what this simulation motivates. The AR bootstrap is restricted to pure autoregressive processes, only. Figure 5.6 visualizes that the AR is no longer able to perform as well as in the purely autoregressive scenario. Contrary, this is a tailor-made scenario for the ARAB. It should be able to capture the autoregressive part of the time series what is expected to substantially improve the performance. Indeed, this is what can be observed in Figure 5.6. The ARAB is able to deliver further improved results in relation to the MBB and is less sensitive to the block length. However, for appropriately chosen block length the MBB still results in satisfying results which shows how strong this approach is even without any further pre-model fits.

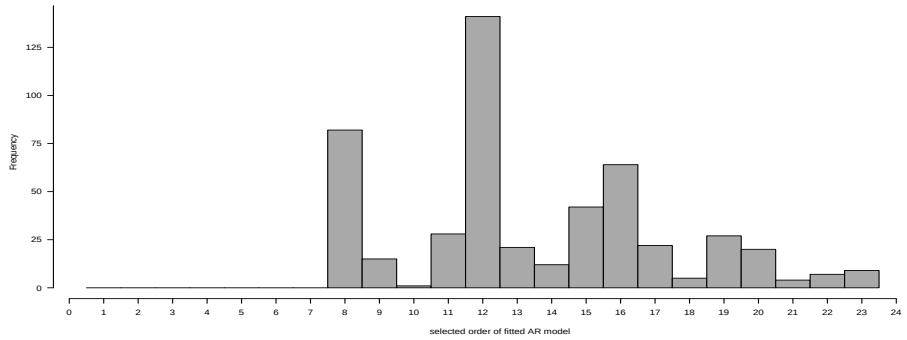

Figure 5.5: ARMA time series: histogram of the selected orders of the AR fits under usage of AIC (absolute frequency of orders out of 500 repetitions is given).

The by AIC chosen block lengths highly fluctuate in this simulation study, cf. Figure 5.5. Anyway, once again it should be emphasized that the block based methods do not necessarily require these model fits. The model fit's use is restricted to improving the ARAB as far as possible. This is independent on the model orders.

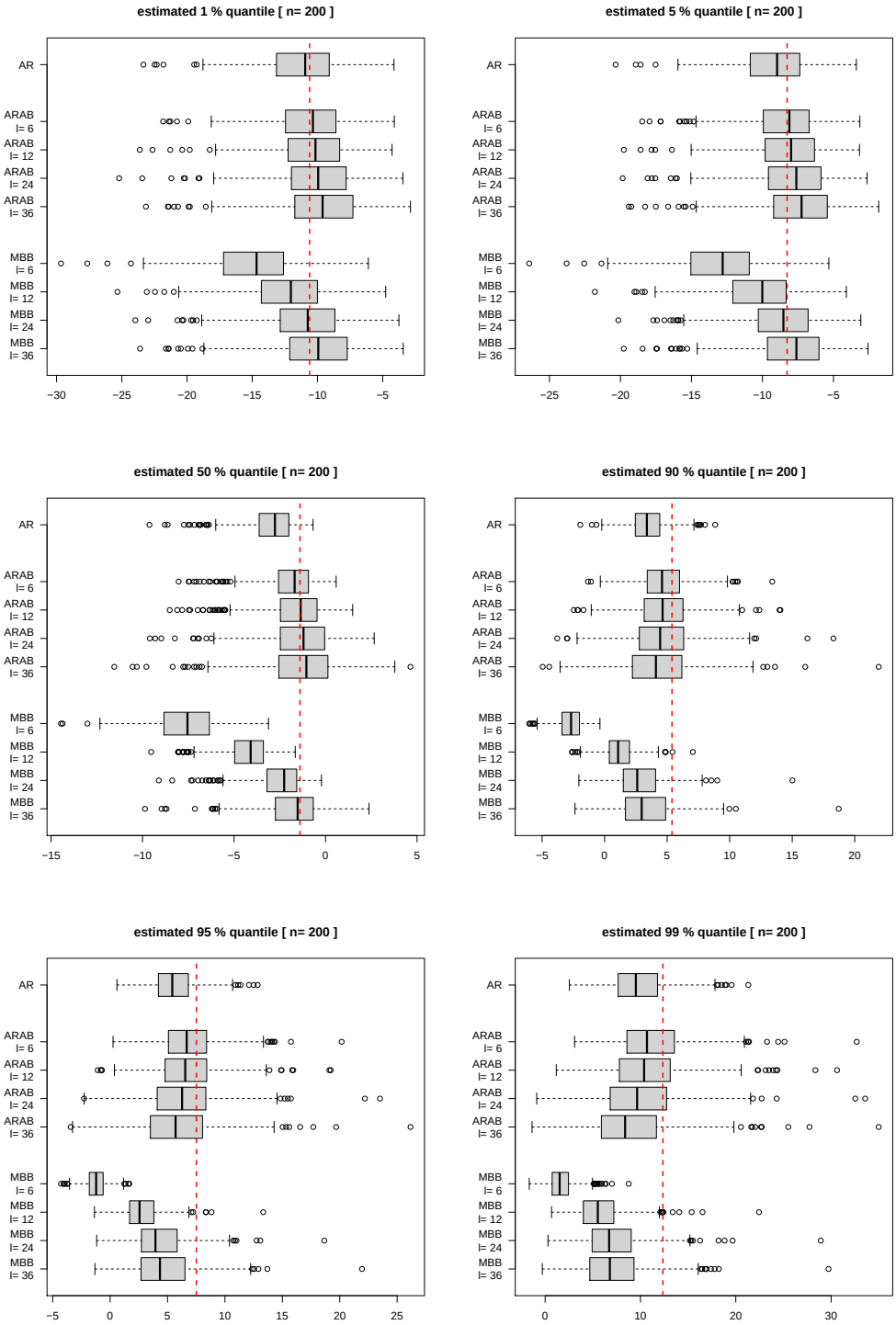

Figure 5.6: ARMA Time Series: boxplots of bootstrapped quantiles of the AR bootstrap, the ARAB bootstrap and MBB bootstrap with true finite sample quantile (dashed red line).

5.5.4 Nonlinear AR Time Series

In contrast to the simulation studies above where either the AR or the ARAB where favoured due to the choice of the considered model, we further consider scenarios where none approach is a priori favoured. In the following, we consider two different nonlinear AR time series.

Nonlinear AR Time Series I.
We revisit the nonlinear autoregressive model as previously used by Paparoditis and Politis (2001) and Shao (2010):

$$X_t = 0.6 \sin(X_{t-1}) + e_t, \quad t \in \mathbb{Z}, \tag{5.29}$$

where $\{e_t : t \in \mathbb{Z}\}$ are i.i.d. $\mathcal{N}(0,1)$. The simulation setup remains as before. Comments on the AR bootstrap are omitted since our focus is on the comparison of the two block-based approaches. The simulation results are visualized in Figures 5.7 and 5.8. Similar to the autoregressive scenario the ARAB is much less sensitive to the choice of the block length than the MBB. The MBB performance is very well for larger block lengths, anyway. In comparison, the ARAB yields the goal distribution also for very short blocks. Summarizing this study the MBB is suitable and well-performing while the ARAB delivers further improved results.

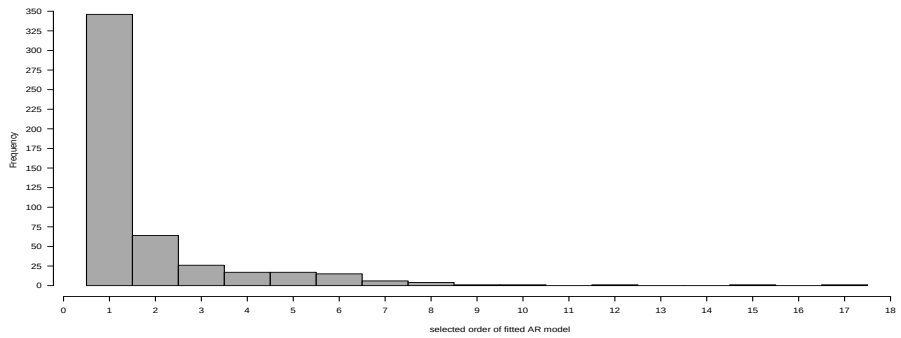

Figure 5.7: Nonlinear AR Time Series I (sin): histogram of the selected orders of the AR fits under usage of AIC (absolute frequency of orders out of 500 repetitions is given).

Nonlinear AR Time Series II.
Another nonlinear time series was considered. The ensuing simulation study was done for the time series

$$X_t = 0.6 \arctan(X_{t-1}) + e_t, \quad t \in \mathbb{Z}, \tag{5.30}$$

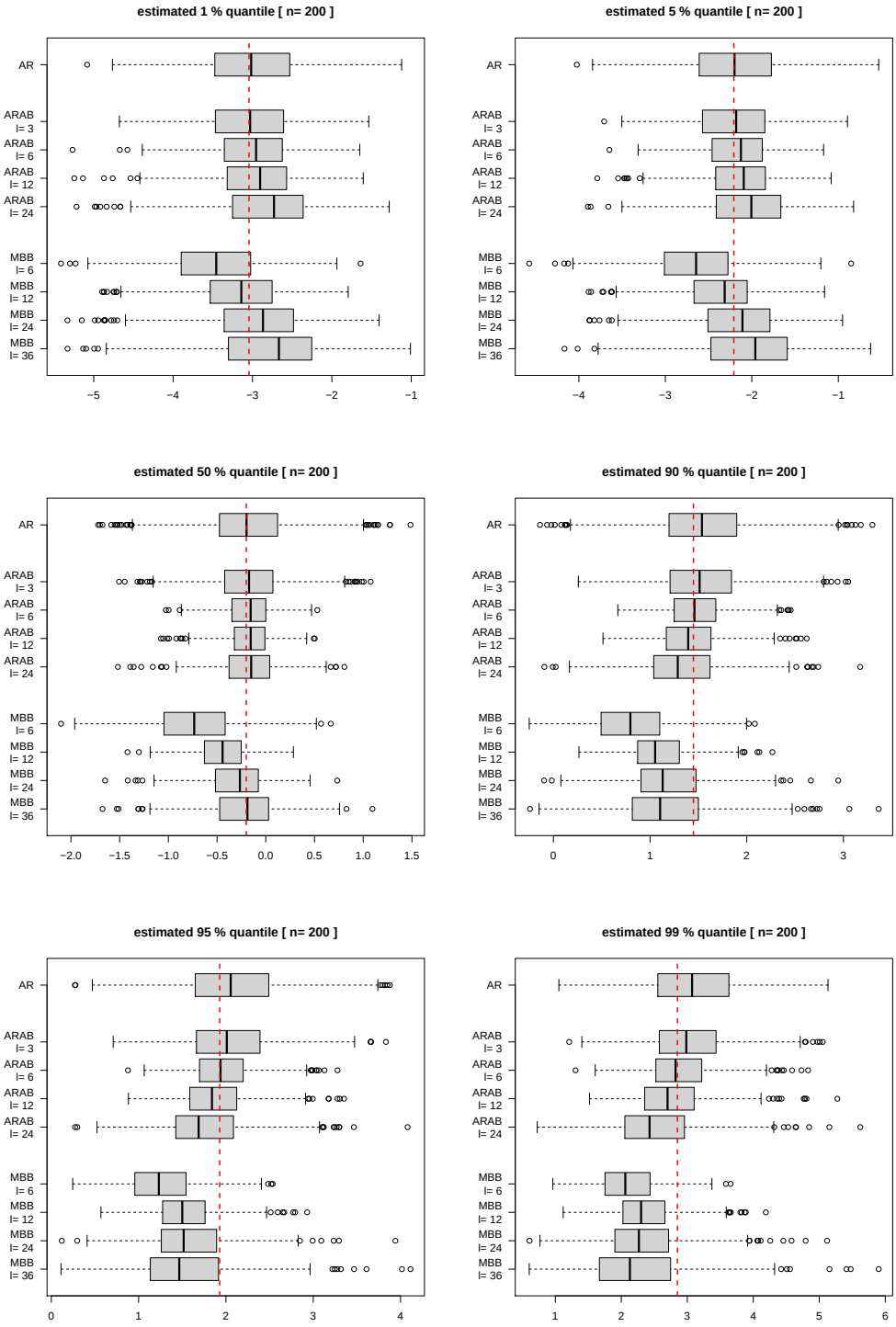

Figure 5.8: Nonlinear AR Time Series I (sin): boxplots of bootstrapped quantiles of the AR bootstrap, the ARAB bootstrap and MBB bootstrap with true finite sample quantile (dashed red line).

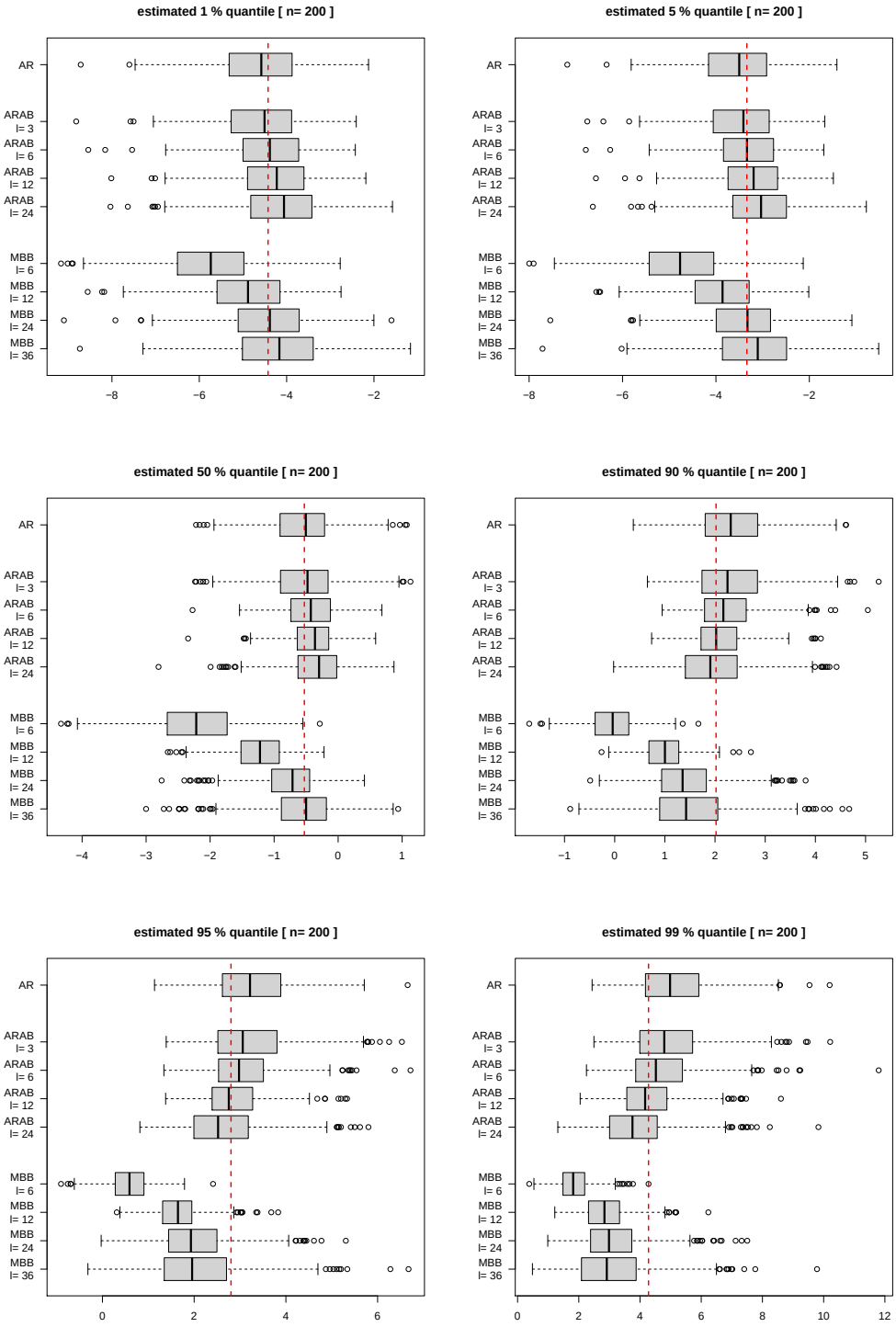

Figure 5.9: Nonlinear AR time series II (arctan): boxplots of bootstrapped quantiles of the AR bootstrap, the ARAB bootstrap and MBB bootstrap with true finite sample quantile (dashed red line).

where $\{e_t : t \in \mathbb{Z}\}$ are i.i.d. $\mathcal{N}(0,1)$. The crucial points and observations coincide with the findings for the nonlinear autoregression with the sin-function above. The findings are summarized in Figures 5.9 and 5.10.

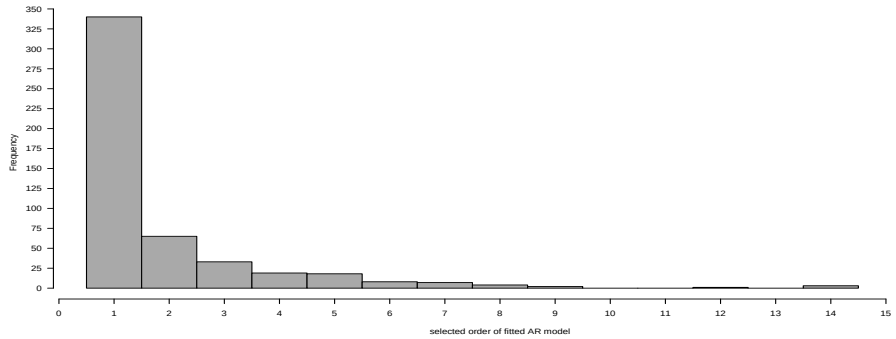

Figure 5.10: Nonlinear AR time series II (arctan): histogram of the selected orders of the AR fits under usage of AIC (absolute frequency of orders out of 500 repetitions is given).

5.5.5 Real world times series

To investigate how the ARAB performs in comparison to the further approaches in real world scenarios, we apply the procedures to real-life data as follows.

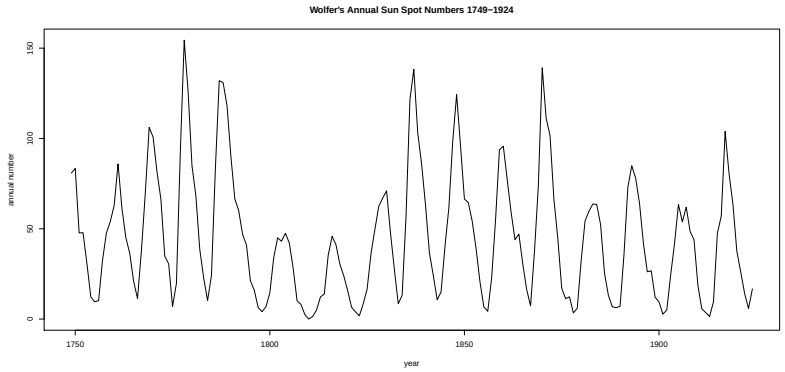

Figure 5.11: Wolfer's Sunspot Numbers from 1749 to 1924.

Wolfer's Sunspot Numbers.
We consider Wolfer's sunspot numbers as given in Anderson (1994). This time series is one of the oldest and most popular time series. It consists of $n = 176$ annual sun spot numbers from 1749 to 1924. As before, we compare the AR to the ARAB and

the MBB. The block sizes are chosen as $l = 6, 12, 24, 36$. Bootstrap 95% confidence intervals are given for the statistic

$$\sqrt{n}(\hat{\varrho}(h) - \varrho(h)), \quad h = 1, 2, 3. \tag{5.31}$$

The time series is visualized in Figure 5.11. The results are visualized in Figure 5.12. Since the time series was often shown to perform as an AR process very likely, it is not surprising that both the AR and the ARAB perform comparable. As before, the MBB shows higher sensitivity to the block length than the ARAB.

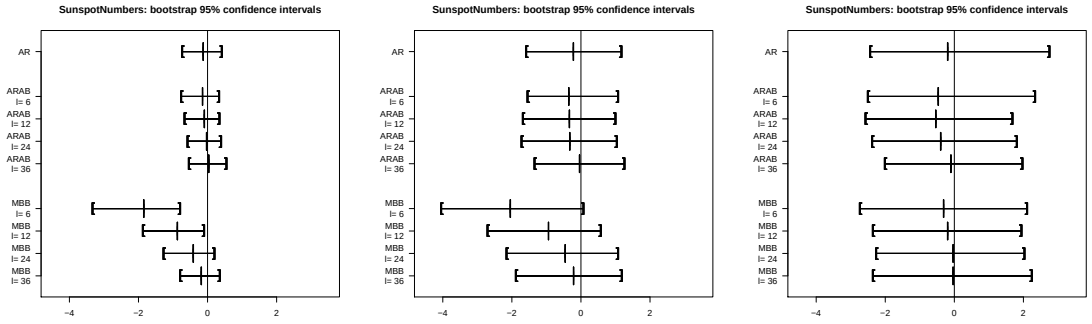

Figure 5.12: Wolfer's Sunpot Numbers: bootstrapped 95% confidence intervals for $\sqrt{n}(\hat{\varrho}(h) - \varrho(h))$, $h = 1, 2, 3$ (from left to right).

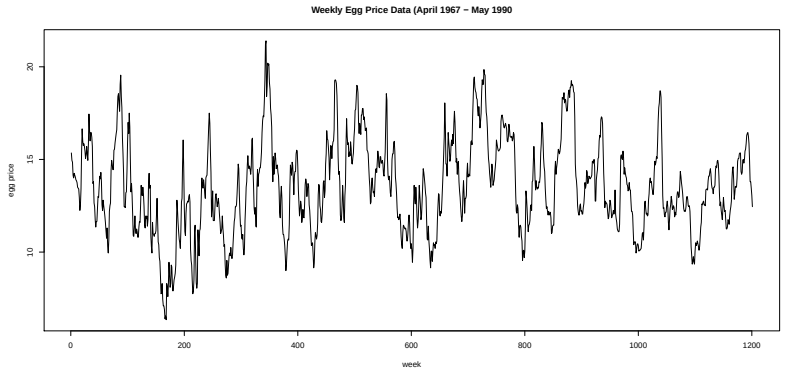

Figure 5.13: Weekly Egg Price Data from at a German agricultural market between April 1967 and May 1990.

Egg Price Data.

We consider the egg-price dataset analyzed in Fan and Yao (2003) and investigated in Neumann and Paparoditis (2008). The dataset consists of $n = 1201$ weekly egg prices

at a German agricultural market between April 1967 and May 1990. Fan and Yao (2003) stated the data exhibits clear nonstationarity, but were able to suitably adapt an ARMA(2,1) model and a MA(7) model to subseries of the dataset. However, in both cases the AR bootstrap should fail. As before, we compare the AR to the ARAB and the MBB. The block sizes are chosen as $l = 50, 100, 180, 250$. Bootstrap 95% confidence intervals are given for the statistic

$$\sqrt{n}(\hat{\varrho}(h) - \varrho(h)), \quad h = 1, 2, 3. \tag{5.32}$$

The results are visualized in Figure 5.14. The ARAB results in more stable intervals than the MBB and its confidence intervals stabilize around zero. Furthermore, for sufficiently large block sizes its variability decreases.

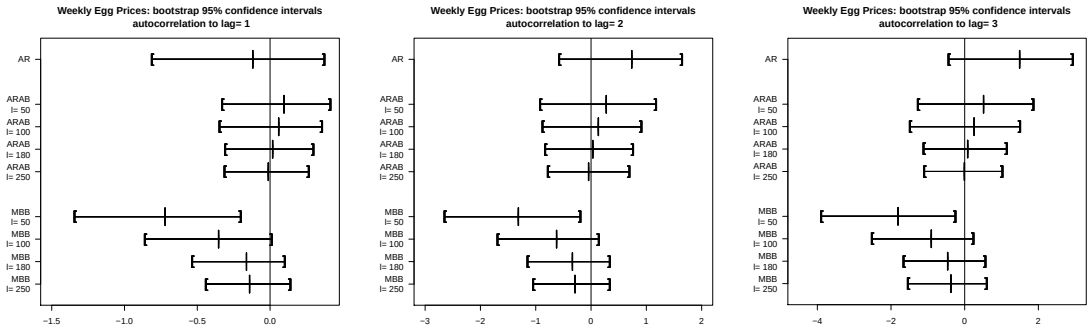

Figure 5.14: Egg price data: Bootstrapped 95% confidence intervals for $\sqrt{n}(\hat{\varrho}(h) - \varrho(h))$, $h = 1, 2, 3$ (from left to right).

5.6 Discussion and conclusions

We have proposed a new resampling method for time series, the autoregressive-aided block bootstrap (ARAB), which is applicable as widely as the moving block bootstrap (MBB) approach. Compared with existing block-based bootstrap methods, the ARAB bootstrap has some charming features:

1. The ARAB was shown to be applicable for the large class of functions of generalized means only requiring some weakly dependent time series.

2. The ARAB combines two theories of resampling techniques, the model-based approach and the block-based approaches. Setting the block length l to 1, it reduces to the ordinary AR bootstrap; setting the fit order to zero one comes

out with the ordinary MBB. Further, setting both the block length to 1 and the fit order to zero, the ARAB reduces to Efron's bootstrap.

3. The ARAB robustifies the AR bootstrap in the sense that misspecification of the underlying model order does not lead to incorrect asymptotics.

4. Although the ARAB uses an AR fit, the underlying model is not required to be of AR form. Any weakly stationary process can be addressed. The AR fit plays the role of a prewhitener and is not the specifier of the model.

5. The ARAB is tailor-made but not limited to weak ARMA time series, low-fequency samples of AR processes, vector-valued AR processes from which only the first component is observed and discretely sampled continuous-time ARMA processes.

<table>
<tr><th colspan="2" rowspan="2">bootstrap approaches</th><th colspan="4">order of AR fit</th></tr>
<tr><th>0</th><th>1</th><th>...</th><th>p_{max}</th></tr>
<tr><td rowspan="4">block length</td><td>1</td><td>Efron</td><td colspan="3">AR</td></tr>
<tr><td>...</td><td rowspan="3">MBB</td><td colspan="3" rowspan="3">ARAB</td></tr>
<tr><td>...</td></tr>
<tr><td>l_{max}</td></tr>
</table>

Figure 5.15: The relation of the ARAB to further bootstrap approaches.

The finite sample performances during the simulations emphasize these advantages and show that the ARAB is much less sensitive to varying block lengths than the MBB.

From our point of view the MBB is a well-established procedure which has shown to perform pretty well in application cases. The ARAB states an extension which shall lead to further improvements.

Especially for time series with AR representation where the innovation sequence cannot be assured to be i.i.d., the ARAB seems tailor-made. These time series arise quite naturally when observing on lower frequencies or from continuous-time processes. However, the MBB is still valid for these processes but does not use the information on the existing AR representation. It seems preferable to make use of this information and combine generality and model accuracy.

Modifications of the ARAB seem possible. So one might think of other model-aided approaches like MA-aided or GARCH-aided block bootstraps. However, in the sense of prewhitening the AR model was often shown to nicely address peaks in the spectral density.

5.7 Proofs and auxiliary results

Lemma 5.9. *Assume that $X_1^*, \ldots, X_n^*$ is a bootstrap sequence generated as described in Section 5.2.2. Further Assumption 4–8 may hold. The ordinary bootstrap time series then has the representation*

$$X_t^* = \sum_{j=0}^{\infty} \hat{\psi}_j U_{t-j}^* \quad \forall t = 1, \ldots, n. \tag{5.33}$$

Define

$$\check{X}_t^* = \sum_{j=0}^{(t-1) \bmod l + M(n)} \hat{\psi}_j U_{t-j}^* \tag{5.34}$$

as a slight variation of X_t^*, where the rate $M(n)$ fulfills $M(n) \leq l$ and $M(n) \to \infty$, as $n \to \infty$. Then, it holds

$$E^* \left| \frac{1}{\sqrt{n-m+1}} \sum_{t=1}^{n-m+1} \left(g\left(\underline{X}_t^*\right) - g\left(\underline{\check{X}}_t^*\right) \right) \right| \to_{n\to\infty} 0 \quad i.p. \tag{5.35}$$

Proof of Lemma 5.9:
Consider $\check{X}_t^*$. This modified version of the bootstrap process is truncated at a time point depending on the time index t. In detail, the truncation is done at a point being $M(n)$ steps in the past of the beginning of the block where the present time point t is located. Thus, the bootstrap observation at time point $t = (r-1)l + s$ is truncated $M(n)$ steps in front of the beginning of the $(r-1)$-th block. Thus, the infinite moving average representation reduces to $s + M(n)$ summands. Since $M(n) \leq l$ the truncated version only correlates with two blocks of residuals. This is advantageous throughout computations, especially when computing correlations or covariances as required in the proof of Theorem 5.1.

It should be explicitly noted that - to the author's knowledge - it is not sufficient to truncate the series at the beginning of the present block only, i.e. to set the truncated series to $\sum_{j=0}^{(t-1) \bmod l}$, to conclude the proof. An exception arises in case of the

sample mean. However, for the general statistic this truncation cannot address the whole intrinsic process structure. The crucial point is that the whole past of each realization has to be included (asymptotically) in the statistic. Hence, as $M(n) \to \infty$ this is sufficient.

Now we come to the proof of the Lemma. Denote $\underline{X}_t^* = (X_t^*, \ldots, X_{t+m-1}^*)$ and $\underline{\check{X}}_t^* = (\check{X}_t^*, \ldots, \check{X}_{t+m-1}^*)$ and define the function

$$s(\underline{x}) := \sum_{i=1}^{q} c_i g_i(\underline{x}), \tag{5.36}$$

for some arbitrary but fixed $\underline{c} = (c_1, \ldots, c_q) \in \mathbb{R}^q$. Now one expands $s\left(\underline{X}_t^*\right)$ at $s\left(\underline{\check{X}}_t^*\right)$ by using the Taylor expansion and obtains

$$\begin{aligned} & s\left(X_t^*, \ldots, X_{t+m-1}^*\right) - s\left(\check{X}_t^*, \ldots, \check{X}_{t+m-1}^*\right) \\ & = \sum_{1 \le |\alpha| \le h-1} \frac{1}{\alpha!} D^\alpha s\left(\underline{\check{X}}_t^*\right) \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha + \sum_{|\alpha|=h} \frac{1}{\alpha!} D^\alpha s\left(\underline{\tau}\right) \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \end{aligned} \tag{5.37}$$

for some $\underline{\tau} = \underline{\check{X}}_t^* + \lambda\left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)$, $\lambda \in [0,1]$, where $\alpha \in \mathbb{N}_0^m$, $|\alpha| = \alpha_1 + \ldots + \alpha_m$, $(\underline{x})^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_m^{\alpha_m}$ and $D^\alpha s(\underline{x}) := \frac{\partial^{|\alpha|} s(\underline{y})}{\partial y_1^{\alpha_1} \cdot \ldots \cdot y_m^{\alpha_m}}|_{\underline{y}=\underline{x}}$. Now, we proceed similarly to Bühlmann (1997). Denote by $\|\cdot\|_r$ the usual $\mathcal{L}_r$-norm with respect to the bootstrap distribution. From (5.37) we then get for

$$\begin{aligned} & \left\| s\left(\underline{X}_t^*\right) - s\left(\underline{\check{X}}_t^*\right) \right\|_1 \\ & \le \sum_{1 \le |\alpha| \le h-1} \frac{E^*\left|D^\alpha s\left(\underline{\check{X}}_t^*\right)\right|}{\alpha!} \left\| \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 + \sum_{|\alpha|=h} \frac{1}{\alpha!} \left\| D^\alpha s(\underline{\tau}) \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 . \end{aligned} \tag{5.38}$$

By Assumption 6 the h-th derivative of $g(\cdot)$, and consequently $s(\cdot)$, is Lipschitz and one obtains for the summands of the second sum

$$\begin{aligned} & \left\| D^\alpha s(\underline{\tau}) \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 \\ & \le E^*\left|D^\alpha s\left(\underline{\check{X}}_t^*\right)\right| \cdot \left\| \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 + \left\| \left(D^\alpha s(\underline{\tau}) - D^\alpha s\left(\underline{\check{X}}_t^*\right)\right) \cdot \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 \\ & \le E^*\left|D^\alpha s\left(\underline{\check{X}}_t^*\right)\right| \cdot \left\| \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 + C_g \lambda \sum_{j=0}^{m-1} \left\| \left(X_{t+j}^* - \check{X}_{t+j}^*\right) \left(\underline{X}_t^* - \underline{\check{X}}_t^*\right)^\alpha \right\|_1 , \end{aligned} \tag{5.39}$$

for some suitable constant C_g, e.g. $C_g := \max_u C_u$. Furthermore, by Assumption 6 and the definitions of $D^\alpha s(\cdot)$ and $\underline{\check{X}}_t^*$, it holds

$$E^*\left|D^\alpha s\left(\underline{\check{X}}_t^*\right)\right| \le E^*\left|D^\alpha s\left(\underline{0}\right)\right| + E^*\left|D^\alpha s\left(\underline{\check{X}}_t^*\right) - D^\alpha s\left(\underline{0}\right)\right| = \mathcal{O}_P(1) \tag{5.40}$$

for any $\alpha \leq h-1$. The same holds for $\underline{\check{X}}_t^*$ replaced by $\underline{X}_t^*$. This leads to

$$
\begin{aligned}
&\left\| \frac{1}{\sqrt{n-m+1}} \sum_{t=1}^{n-m+1} \left(s\left(\underline{X}_t^*\right) - s\left(\underline{\check{X}}_t^*\right)\right) \right\|_1 \qquad (5.41)\\
&\leq \frac{1}{\sqrt{n-m+1}} \sum_{i=1}^{b} \sum_{k=1}^{l-m+1} \left(\sum_{1\leq|\alpha|\leq h-1} \frac{E^*\left|D^\alpha s\left(\underline{\check{X}}_{(i-1)l+k}^*\right)\right|}{\alpha!} \right.\\
&\qquad \left. \cdot \left\| \left(\underline{X}_{(i-1)l+k}^* - \underline{\check{X}}_{(i-1)l+k}^*\right)^\alpha \right\|_1 \right.\\
&\qquad \left. + \sum_{|\alpha|=h} \frac{1}{\alpha!} \left\| D^\alpha s(\underline{\tau}) \left(\underline{X}_{(i-1)l+k}^* - \underline{\check{X}}_{(i-1)l+k}^*\right)^\alpha \right\|_1 \right) + o_P(1)\\
&\leq \frac{1}{\sqrt{n-m+1}} \sum_{i=1}^{b} \sum_{k=1}^{l-m+1} \left(\sum_{1\leq|\alpha|\leq h-1} \frac{E^*\left|D^\alpha s\left(\underline{\check{X}}_{(i-1)l+k}^*\right)\right|}{\alpha!} \right.\\
&\qquad \left. \cdot \left\| \left(\sum_{j=((i-1)l+k) \bmod l+M(n)+1}^{\infty} \hat{\psi}_j \underline{U}_{(i-1)l+k-j}^* \right)^\alpha \right\|_1 \right.\\
&\qquad \left. + \sum_{|\alpha|=h} \frac{1}{\alpha!} E^*\left|D^\alpha s\left(\underline{\check{X}}_{(i-1)l+k+v}^*\right)\right| \cdot \left\| \left(\underline{X}_{(i-1)l+k+v}^* - \underline{\check{X}}_{(i-1)l+k+v}^*\right)^\alpha \right\|_1 \right.\\
&\qquad \left. + \sum_{|\alpha|=h} \frac{1}{\alpha!} C_g \lambda \sum_{v=0}^{m-1} \left\| \left(X_{(i-1)l+k+v}^* - \check{X}_{(i-1)l+k+v}^*\right) \right.\right.\\
&\qquad \left.\left. \cdot \left(\underline{X}_{(i-1)l+k}^* - \underline{\check{X}}_{(i-1)l+k}^*\right)^\alpha \right\|_1 \right) + o_P(1).
\end{aligned}
$$

Now we seperately consider the terms in matrix norm. The modulo condition can be simplified since $((i-1)l+k) \bmod l + M(n) + 1 = k + M(n) + 1$. Furthermore, it is well-known (e.g. Kreiss and Neuhaus (2006)) that the coefficients ψ_j (and $\hat{\psi}_j$) of the inverse autoregressive polynomial (and its estimated counterpart) uniformly yield

$$
|\psi_j| \leq \rho^j, \quad \forall j \in \mathbb{Z}, \tag{5.42}
$$

for some $0 < \rho < 1$ which depends on the autoregressive parameters. Thus, by using Assumption 7 one computes

$$
\begin{aligned}
&\left\| \left(\sum_{j=k+M(n)+1}^{\infty} \hat{\psi}_j \underline{U}_{(i-1)l+k-j}^* \right)^\alpha \right\|_1 \qquad (5.43)\\
&= E^* \left| \left(\sum_{j=k+M(n)+1}^{\infty} \hat{\psi}_j U_{(i-1)l+k-j}^* \right)^{\alpha_1} \cdot \ldots \cdot \left(\sum_{j=k+M(n)+1}^{\infty} \hat{\psi}_j U_{(i-1)l+k+m-1-j}^* \right)^{\alpha_m} \right|\\
&\leq \prod_{r=1}^{m} E^* \left| \left(\sum_{j=k+M(n)+1}^{\infty} \hat{\psi}_j U_{(i-1)l+k-j}^* \right)^{\alpha_r} \right|
\end{aligned}
$$

$$
\begin{aligned}
&= \prod_{r=1}^{m} E^* \left| \sum_{j_1,\ldots,j_{\alpha_r}=k+M(n)+1}^{\infty} \hat{\psi}_{j_1} \cdot \ldots \cdot \hat{\psi}_{j_{\alpha_r}} U^*_{(i-1)l+k-j_1} \cdot \ldots \cdot U^*_{(i-1)l+k-j_{\alpha_r}} \right| \\
&\leq \prod_{r=1}^{m} \sum_{j_1,\ldots,j_{\alpha_r}=k+M(n)+1}^{\infty} \rho^{j_1} \cdot \ldots \cdot \rho^{j_{\alpha_r}} E^* \left| U^*_{(i-1)l+k-j_1} \cdot \ldots \cdot U^*_{(i-1)l+k-j_{\alpha_r}} \right| \\
&= \prod_{r=1}^{m} \sum_{j_1,\ldots,j_{\alpha_r}=0}^{\infty} \rho^{k+M(n)+1+j_1} \cdot \ldots \cdot \rho^{k+M(n)+1+j_{\alpha_r}} \cdot \mathcal{O}_P(1) \\
&= \rho^{(k+M(n)+1)|\alpha|} \prod_{r=1}^{m} \sum_{j_1,\ldots,j_{\alpha_r}=0}^{\infty} \rho^{j_1} \cdot \ldots \cdot \rho^{j_{\alpha_r}} \cdot \mathcal{O}_P(1) \\
&= \rho^{(k+M(n)+1)|\alpha|} \cdot C' \cdot \mathcal{O}_P(1),
\end{aligned}
$$

where C' yields the obvious notation. The other terms in matrix norm can be handled in direct analogy. The second term yields exactly the same result while for the third term the constant C' has to be replaced by $C'' = C' \cdot \sum_{j=0}^{\infty} \rho^j$ and $|\alpha|$ by $|\alpha| + 1$. Hence,

$$
\begin{aligned}
&\left\| \left(X^*_{(i-1)l+k+v} - \check{X}^*_{(i-1)l+k+v} \right) \left(\underline{X}^*_{(i-1)l+k} - \underline{\check{X}}^*_{(i-1)l+k} \right)^{\alpha} \right\|_1 \\
&= \rho^{(k+M(n)+1)(|\alpha|+1)} \cdot C'' \cdot \mathcal{O}_P(1).
\end{aligned} \tag{5.44}
$$

Altogether it holds

$$
\begin{aligned}
&\left\| \frac{1}{\sqrt{n-m+1}} \sum_{t=1}^{n-m+1} \left(s\left(\underline{X}^*_t\right) - s\left(\underline{\check{X}}^*_t\right) \right) \right\|_1 \\
&\leq \frac{1}{\sqrt{n-m+1}} \sum_{i=1}^{b} \sum_{k=1}^{l-m+1} \Bigg(\sum_{1 \leq |\alpha| \leq h-1} \frac{1}{\alpha!} \rho^{(k+M(n)+1)|\alpha|} \cdot C' \cdot \mathcal{O}_P(1) \\
&\qquad + \sum_{|\alpha|=h} \frac{1}{\alpha!} \rho^{(k+M(n)+1)|\alpha|} \cdot C' \cdot \mathcal{O}_P(1) \\
&\qquad + \sum_{|\alpha|=h} \frac{1}{\alpha!} C_g \lambda \sum_{v=0}^{m-1} \rho^{(k+M(n)+1)(|\alpha|+1)} \cdot C'' \cdot \mathcal{O}_P(1) \Bigg) \\
&\qquad + o_P(1) \\
&= \frac{\mathcal{O}_P(1)}{\sqrt{n-m+1}} \, b \, \rho^{M(n)} \Bigg(\sum_{1 \leq |\alpha| \leq h-1} \rho^{|\alpha|} \sum_{k=1}^{l-m+1} \rho^{k|\alpha|} \\
&\qquad + \sum_{|\alpha|=h} \rho^{|\alpha|} \sum_{k=1}^{l-m+1} \rho^{k|\alpha|} + \sum_{|\alpha|=h} \rho^{|\alpha|+1} \sum_{k=1}^{l-m+1} \rho^{k(|\alpha|+1)} \Bigg) \\
&= \mathcal{O}_P\left(\frac{b}{\sqrt{n}} \, \rho^{M(n)} \right)
\end{aligned} \tag{5.45}
$$

which converges to zero in probability as long as $M(n) \to \infty$ and concludes the proof of the lemma. □

Proof of Theorem 5.1:
Having Lemma 5.9 at hand, it suffices to show the theorem for T_n^* with X_t^* replaced by $\check{X}_t^*$. For the truncated series choose $M(n)$ such that $M(n)^2 l^{-1} \to 0$ as $n \to \infty$. Consider first part *(i)*, the limiting covariance matrix. As discussed at the beginning of the proof of Lemma 5.9, each $\check{X}_t^*$ only is correlated with at least two blocks of residuals. Thus, two bootstrap realizations $\check{X}_t^*$ and $\check{X}_s^*$ either stem from the same block, or $\check{X}_t^*$ stems from the block before or after $\check{X}_s^*$, or the realizations are uncorrelated. Hence,

$$
\begin{aligned}
& Cov^*\left(\frac{1}{\sqrt{n-m+1}}\sum_{t=1}^{n-m+1} g_u\left(\underline{\check{X}}_t^*\right), \frac{1}{\sqrt{n-m+1}}\sum_{s=1}^{n-m+1} g_v\left(\underline{\check{X}}_s^*\right)\right) \qquad (5.46)\\
&= \frac{1}{n-m+1} Cov^*\left(\sum_{i_1=1}^{b}\sum_{k_1=1}^{l-m+1} g_u\left(\underline{\check{X}}_{(i_1-1)l+k_1}^*\right), \sum_{i_2=1}^{b}\sum_{k_2=1}^{l-m+1} g_v\left(\underline{\check{X}}_{(i_2-1)l+k_2}^*\right)\right)\\
&\quad + o_P(1)\\
&= \frac{1}{n-m+1}\sum_{i=1}^{b}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{(i-1)l+k_1}^*\right), g_v\left(\underline{\check{X}}_{(i-1)l+k_2}^*\right)\right)\\
&\qquad + \frac{1}{n-m+1}\sum_{i=1}^{b-1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{il+k_1}^*\right), g_v\left(\underline{\check{X}}_{(i-1)l+k_2}^*\right)\right)\\
&\qquad + \frac{1}{n-m+1}\sum_{i=1}^{b-1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{(i-1)l+k_1}^*\right), g_v\left(\underline{\check{X}}_{il+k_2}^*\right)\right) + o_P(1)\\
&= \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{k_1}^*\right), g_v\left(\underline{\check{X}}_{k_2}^*\right)\right)\\
&\qquad + \frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{l+k_1}^*\right), g_v\left(\underline{\check{X}}_{k_2}^*\right)\right)\\
&\qquad + \frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov^*\left(g_u\left(\underline{\check{X}}_{k_1}^*\right), g_v\left(\underline{\check{X}}_{l+k_2}^*\right)\right) + o_P(1)\\
&= \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov\left(g_u\left(\underline{X}_{k_1}\right), g_v\left(\underline{X}_{k_2}\right)\right)\\
&\qquad + \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right) - g_u\left(\underline{X}_{k_1}\right), g_v\left(\underline{X}_{k_2}\right)\right)\\
&\qquad + \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1} Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right) - g_v\left(\underline{X}_{k_2}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}\Big(\,Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right),g_v\left(\underline{\check{X}}^*_{k_2}\right)\right)\\
&\qquad\qquad -Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right),g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{X}_{l+k_1}\right),g_v\left(\underline{X}_{k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{\check{X}}_{l+k_1}\right)-g_u\left(\underline{X}_{l+k_1}\right),g_v\left(\underline{X}_{k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{\check{X}}_{l+k_1}\right),g_v\left(\underline{\check{X}}_{k_2}\right)-g_v\left(\underline{X}_{k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}\Big(\,Cov^*\left(g_u\left(\underline{\check{X}}^*_{l+k_1}\right),g_v\left(\underline{\check{X}}^*_{k_2}\right)\right)\\
&\qquad\qquad -Cov\left(g_u\left(\underline{\check{X}}_{l+k_1}\right),g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{X}_{k_1}\right),g_v\left(\underline{X}_{l+k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right)-g_u\left(\underline{X}_{k_1}\right),g_v\left(\underline{X}_{l+k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right),g_v\left(\underline{\check{X}}_{l+k_2}\right)-g_v\left(\underline{X}_{l+k_2}\right)\right)\\
&+\frac{b-1}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}\Big(\,Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right),g_v\left(\underline{\check{X}}^*_{l+k_2}\right)\right)\\
&\qquad\qquad -Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right),g_v\left(\underline{\check{X}}_{l+k_2}\right)\right)\Big)\\
&+o_P(1)\\
=:&\sum_{i=1}^{12}R_i+o_P(1),
\end{aligned}
$$

where the expressions $R_1,\ldots,R_{12}$ yield the obvious definition. In the following, we show that $R_1+R_5+R_9$ converges to the limiting convariance of interest, namely $(\Sigma_{q\times q})_{u,v}$. All other remainder terms are shown to converge to zero. Special interest is on the expression R_4 (as well as R_8 and R_{12}) which validates that the bootstrap covariance converges to its counterpart for the original time series. However, we start with the computation for $R_1+R_5+R_9$. By using formulae (A.11) of Kreiss and Neuhaus (2006) and direct computation, one obtains

$$
R_1 \;=\; \frac{b}{n-m+1}\sum_{h=-(l-m)}^{l-m}\left((l-m+1)-|h|\right)Cov(g_u(\underline{X}_0),g_v(\underline{X}_h)) \quad (5.47)
$$

$$
\begin{aligned}
R_5 &= \frac{b-1}{n-m+1}\sum_{h=m}^{l}(h-m+1)\,Cov(g_u(\underline{X}_h),g_v(\underline{X}_0)) \qquad (5.48)\\
&\quad+\frac{b-1}{n-m+1}\sum_{h=l+1}^{2l-m}(2l-m+1-h)\,Cov(g_u(\underline{X}_h),g_v(\underline{X}_0))\\
R_9 &= \frac{b-1}{n-m+1}\sum_{h=m}^{l}(h-m+1)\,Cov(g_u(\underline{X}_0),g_v(\underline{X}_h)) \qquad (5.49)\\
&\quad+\frac{b-1}{n-m+1}\sum_{h=l+1}^{2l-m}(2l-m+1-h)\,Cov(g_u(\underline{X}_0),g_v(\underline{X}_h))
\end{aligned}
$$

what immediately leads to the aggregation

$$
\begin{aligned}
&R_1+R_5+R_9 \qquad (5.50)\\
&=\frac{b}{n}\Bigg[\sum_{h=-l}^{l} l\,Cov(g_u(\underline{X}_0),g_v(\underline{X}_h))\\
&\qquad+\sum_{h=l+1}^{2l-m}(2l-m+2-h)\,Cov(g_u(\underline{X}_0),g_v(\underline{X}_h))\\
&\qquad+\sum_{h=-(l+1)}^{-(2l-m)}(2l-m+1-|h|)\,Cov(g_u(\underline{X}_0),g_v(\underline{X}_h))\Bigg]+o(1)\\
&\to_{n\to\infty}\sum_{h=-\infty}^{\infty}Cov(g_u(\underline{X}_0),g_v(\underline{X}_h))=\left(\Sigma_{q\times q}\right)_{u,v},
\end{aligned}
$$

and shows the convergency to the limiting covariance matrix as given in the Theorem. Now we proceed to proof that R_i converges to zero for all $i=2,3,6,7,10,11$. The strategy is very similar for all these remainder terms. Exemplarily, we consider R_2 in detail. Applying the Taylor expansion to the first argument in R_2 and afterwards using Cauchy Schwarz inequality results to

$$
\begin{aligned}
R_2 &= \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}Cov\Bigg(\sum_{1\le|\alpha|\le h-1}\frac{D^\alpha g_u(\underline{X}_{k_1})}{\alpha!}\left(\check{\underline{X}}_{k_1}-\underline{X}_{k_1}\right)^\alpha \qquad (5.51)\\
&\qquad\qquad+\sum_{|\alpha|=h}\frac{D^\alpha g_u(\underline{\tau})}{\alpha!}\left(\check{\underline{X}}_{k_1}-\underline{X}_{k_1}\right)^\alpha,g_v\left(\underline{X}_{k_2}\right)\Bigg)\\
&\le \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}E\Bigg[\Bigg(\frac{D^\alpha g_u(\underline{X}_{k_1})}{\alpha!}\left(\sum_{j=k_1+M(n)+1}^{\infty}\psi_j U_{k_1-j}\right)^\alpha\\
&\qquad\qquad+\sum_{|\alpha|=h}\frac{D^\alpha g_u(\underline{\tau})}{\alpha!}\left(\check{\underline{X}}_{k_1}-\underline{X}_{k_1}\right)^\alpha\Bigg)^2\Bigg]^{1/2}
\end{aligned}
$$

$$\cdot E\left[g_v\left(\underline{X}_{k_2}\right)^2\right]^{1/2}$$

The first expectation is of same form as the terms investigated in (5.43)-(5.45). Thus, Assumption 7 leads to

$$\begin{aligned} |R_2| &\leq \frac{b}{n-m+1} \sum_{k_1=1}^{l-m+1} \rho^{(k_1+M(n)+1)\cdot 2|\alpha|} \sum_{k_2=1}^{l-m+1} \sqrt{E\left[g_v\left(\underline{X}_{k_2}\right)^2\right]} \qquad (5.52) \\ &\leq \frac{b}{n-m+1} \cdot C_1 \cdot \rho^{M(n)} \cdot (l-m+1) \cdot C_2 \\ &= \mathcal{O}_P\left(\rho^{M(n)}\right), \end{aligned}$$

for suitable constants C_1, C_2, which converges to zero as $n \to \infty$. The other remainder terms R_i, $i = 3, 6, 7, 10, 11$, behave in an analogue way and since these computations give no further insight, they are omitted here.

Hence, it only remains to investigate R_4, R_8 and R_{12} to conclude the computation of the limiting covariance matrix and to proof this part of the proof. To show that R_4 asymptotically vanishes in probability, we mainly follow the lines of the proof of Lemma 5.5 in Bühlmann (1997). Thus, approach

$$\begin{aligned} \check{X}_t^* &= \sum_{j=0}^{(t-1) \bmod l + M(n)} \hat{\psi}_j U_{t-j}^* \qquad (5.53) \\ &= \sum_{j=0}^{M(n)} \psi_j U_{t-j}^* + \sum_{j=0}^{M(n)} (\hat{\psi}_j - \psi_j) U_{t-j}^* + \sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \hat{\psi}_j U_{t-j}^* \end{aligned}$$

Now, let $x \in \mathbb{R}$ be a continuity point of the distribution function of X_t. Furthermore, let $\gamma > 0$ be arbitrary. Then using the same proposal as in the proof of Slutsky's Theorem

$$\begin{aligned} &P^*\left[\check{X}_t^* \leq x\right] \qquad (5.54) \\ &\leq P^*\left[\sum_{j=0}^{M(n)} \psi_j U_{t-j}^* \leq x + \gamma\right] \\ &\quad + P^*\left[\left|\sum_{j=0}^{M(n)} \left(\hat{\psi}_j - \psi_j\right) U_{t-j}^*\right| > \frac{\gamma}{2}\right] + P^*\left[\left|\sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \hat{\psi}_j U_{t-j}^*\right| > \frac{\gamma}{2}\right] \end{aligned}$$

and by the Markov inequality

$$P^*\left[\left|\sum_{j=0}^{M(n)} \left(\hat{\psi}_j - \psi_j\right) U_{t-j}^*\right| > \frac{\gamma}{2}\right] \leq 2 \sum_{j=0}^{M(n)} \left|\hat{\psi}_j - \psi_j\right| E^*\left|U_{t-j}^*\right| \frac{1}{\gamma} \qquad (5.55)$$

and

$$P^*\left[\left|\sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \hat{\psi}_j U_{t-j}^*\right| > \frac{\gamma}{2}\right] \leq 2 \sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \left|\hat{\psi}_j\right| E^*\left|U_{t-j}^*\right| \frac{1}{\gamma}. \tag{5.56}$$

By Lemma 2.2, Kreiss and Franke (1992) and as $M(n)$ is sufficiently large we have

$$\sum_{j=0}^{M(n)} \left|\hat{\psi}_j - \psi_j\right| \leq \xi \tag{5.57}$$

for any $\xi > 0$. Furthermore by (5.42) it holds

$$\sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \left|\hat{\psi}_j\right| \leq C\rho^{M(n)}, \tag{5.58}$$

for a suitable constant C. Because of $E^*\left|U_{t-j}^*\right| \leq \sqrt{E^*(U_{t-j}^*)^2}$ and Assumption 7, altogether one obtains, as n is sufficiently large and for any $\kappa > 0$,

$$P^*\left[\left|\sum_{j=0}^{M(n)} \left(\hat{\psi}_j - \psi_j\right) U_{t-j}^*\right| > \frac{\gamma}{2}\right] \leq \frac{\kappa}{2} \quad i.p. \tag{5.59}$$

and

$$P^*\left[\left|\sum_{j=M(n)+1}^{M(n)+(t-1) \bmod l} \hat{\psi}_j U_{t-j}^*\right| > \frac{\gamma}{2}\right] \leq \frac{\kappa}{2} \quad i.p.. \tag{5.60}$$

Therefore

$$P^*\left[\check{X}_t^* \leq x\right] \leq P^*\left[\sum_{j=0}^{M(n)} \psi_j U_{t-j}^* \leq x + \gamma\right] + \kappa \quad i.p. \tag{5.61}$$

and in direct analogy

$$P^*\left[\check{X}_t^* \leq x\right] \geq P^*\left[\sum_{j=0}^{M(n)} \psi_j U_{t-j}^* \leq x - \gamma\right] - \kappa \quad i.p.. \tag{5.62}$$

By the consistency of the Yule-Walker estimates and Assumption 7 one immediately has the consistency of the bootstrap moments

$$E^*\left[(U_t^*)^\alpha\right] \;=\; \frac{1}{n}\sum_{t=1}^{n} \hat{U}_t^\alpha + o_P(1) \xrightarrow{\mathcal{P}} E\left[U_t^\alpha\right] \tag{5.63}$$

for all $\alpha \leq 2(h+1)$. Now we show

$$U_t^* \xrightarrow{\mathcal{D}^*} U_t \;\; i.p.. \tag{5.64}$$

Denote by F_U the distribution function of the process $\{U_t : t \in \mathbb{Z}\}$ and by $F_{\hat{U}}$ the distribution function of its estimated counterpart. Further denote by $\hat{F}_{\hat{U}}$ the empirical version of $F_{\hat{U}}$. To proof (5.64) consider the Mallows metric $d_2(\cdot,\cdot)$ via

$$d_2\left(\hat{F}_{\hat{U}}, F_U\right) \leq d_2\left(\hat{F}_{\hat{U}}, F_{\hat{U}}\right) + d_2\left(F_{\hat{U}}, F_U\right) \tag{5.65}$$

This allows for handling the two summands on the right hand side seperately. The first term converges to zero by the approach as in the proof of Lemma 5.4 of Bühlmann (1997), and using (5.63). The second term converges to zero by Lemma 8.4 of Bickel and Freedman (1981). Together this validates (5.64) and further delivers

$$\left(U^*_{(i-1)l+1}, \ldots, U^*_{(i-1)l+l}\right) \xrightarrow{\mathcal{D}^*} (U_1, \ldots, U_l) \quad i.p. \tag{5.66}$$

for each block of bootstrap innovations. Define the set

$$T_n := \{k \,|\, M(n) + 1 \leq k \leq l\}\,. \tag{5.67}$$

and since $1 \geq M(n)/l \to 0$ we have in probability for any $\kappa > 0$ as n is sufficiently large

$$\left| P^*\left[\sum_{j=0}^{M(n)} \psi_j U^*_{t-j} \leq x + \gamma\right] - P\left[\sum_{j=0}^{M(n)} \psi_j U_{t-j} \leq x + \gamma\right]\right| \leq \kappa, \tag{5.68}$$

for all $t \bmod l \in T_n$. This is a rather relevant point within the proof. As discussed at the beginning of the proof of Lemma (5.9), it is important to assure for $M(n) \to \infty$ to incorporate completely the process' structure. However, (5.68) only can be validated if the bootstrap innovations stem from one single block and hence, if the (truncated) series does not overlap from one block of innovations to another. Otherwise the corresponding non-bootstrap analogue would not be mimicked correctly since the dependence structure is destroyed at the cut between the independent blocks of innovations. Indeed, this is the reason why the set T_n is introduced and furthermore, why the rate $M(n)$ has to increase slower than the block length l. The ensuing computations will clarify more detailed why $M(n)^2 l^{-1} \to 0$ has to be fulfilled. Thus, it holds

$$\check{X}^*_t \xrightarrow{\mathcal{D}^*} \check{X}_t \;\; \forall t \bmod l \in T_n, \tag{5.69}$$

and by the Cramér-Wold device one immediately has, for some arbitrary $d \in \mathbb{N}$, and for any $t_1, \ldots, t_d \in T_n$,

$$\left(\check{X}^*_{t_1}, \ldots, \check{X}^*_{t_d}\right) \xrightarrow{\mathcal{D}^*} \left(\check{X}_{t_1}, \ldots, \check{X}_{t_d}\right) \;\; \forall t \bmod l \in T_n. \tag{5.70}$$

It should be emphasized in here that (5.66) holds blockwise only, and thus, by the above proof, it only validates (5.70) for bootstrap realizations stemming from the set T_n. Bootstrap realizations as truncated above stemming from earlier time points of a block (namely the first $M(n)$ positions) overlap to the previous block of innovations. However, this characteristic is asymptotically negligible as it will be shown lateron.

As a next step we investigate the function g and its component functions. Truncate g_u via

$$\tilde{g}_u(x) = g_u(x)\mathbb{1}_{|g_u(x)|\leq K}(x) + K\, sign(g_u(x))\mathbb{1}_{|g_u(x)|>K}, \quad K > 0. \tag{5.71}$$

Then $\tilde{g}_u(\cdot)$ is continuous and bounded and one immediately has

$$\begin{aligned} &Cov^*\left(\tilde{g}_u\left(\underline{\check{X}}_t^*\right), \tilde{g}_v\left(\underline{\check{X}}_s^*\right)\right) \\ &= Cov\left(\tilde{g}_u\left(\underline{\check{X}}_t\right), \tilde{g}_v\left(\underline{\check{X}}_s\right)\right) + o_P(1) \;\; \forall t, s \bmod l \in T_n \end{aligned} \tag{5.72}$$

by (5.70). Now we show that the effect of truncating the functions is asymptotically negligible. By Hölder's inequality

$$\begin{aligned} &E^*\left|g_u\left(\underline{\check{X}}_t\right)\mathbb{1}_{|g_u(x)|>K}\right|^2 \\ &\leq \left(E^*\left|g_u\left(\underline{\check{X}}_t\right)\right|^{2(h+2)/(h+1)}\right)^{(h+1)/(h+2)} \left(P^*\left(\left|g_u\left(\underline{\check{X}}_t\right)\right| > K\right)\right)^{1/(h+2)} \\ &= \mathcal{O}_P(1)K^{-2/(h+1)}. \end{aligned} \tag{5.73}$$

To see that the expected value is $\mathcal{O}_P(1)$ approach as in the proof of Lemma 5.9. Then, for arbitrary $\kappa > 0$, we can choose $K = K(\kappa, n)$ such that for n sufficiently large and in probability, for all $t, s \bmod l \in T_n$,

$$\left| Cov^*\left(\tilde{g}_u\left(\underline{\check{X}}_t^*\right), \tilde{g}_v\left(\underline{\check{X}}_s^*\right)\right) - Cov^*\left(g_u\left(\underline{\check{X}}_t^*\right), g_v\left(\underline{\check{X}}_s^*\right)\right)\right| \leq \frac{\kappa}{N(n)} \tag{5.74}$$

and in complete analogy, also for all $t, s \bmod l \in T_n$,

$$\left| Cov\left(\tilde{g}_u\left(\underline{\check{X}}_t\right), \tilde{g}_v\left(\underline{\check{X}}_s\right)\right) - Cov\left(g_u\left(\underline{\check{X}}_t\right), g_v\left(\underline{\check{X}}_s\right)\right)\right| \leq \frac{\kappa}{N(n)} \tag{5.75}$$

where $N(n) \to \infty$, as $n \to \infty$, is chosen such that $l/N(n) \to 0$. Hence, it follows, for all $t, s \bmod l \in T_n$,

$$Cov^*\left(g_u\left(\underline{\check{X}}_t^*\right), g_v\left(\underline{\check{X}}_s^*\right)\right) - Cov\left(g_u\left(\underline{\check{X}}_t\right), g_v\left(\underline{\check{X}}_s\right)\right) \tag{5.76}$$

$$= \mathcal{O}_P\left(N(n)^{-1}\right).$$

This result will be helpful in the following. Nevertheless, it does not include covariances having its arguments at the beginning of a block. Therefore, seperately consider for $k_1 \notin T_n$ and $k_2 \in T_n$ and by using (5.76)

$$\begin{aligned} & Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right) \\ = \; & Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) + Cov^*\left(g_u\left(\underline{\check{X}}^*_{l-m+1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ & - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right) \\ = \; & Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ & - Cov\left(g_u\left(\underline{\check{X}}_{k_1} - \underline{\check{X}}_{l-m+1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right) + \mathcal{O}_P(N(n)^{-1}) \end{aligned} \tag{5.77}$$

Then by application of the Taylor expansion obtain for the first covariance

$$\begin{aligned} & Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ \leq \; & \left(E^*\left[\left(\sum_{1\leq|\alpha|\leq h-1} \frac{1}{\alpha!} D^\alpha g_u\left(\underline{\check{X}}^*_{l-m+1}\right)\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right)^\alpha \right.\right.\right. \\ & \left.\left.\left. + \sum_{|\alpha|=h} \frac{1}{\alpha!} D^\alpha g_u\left(\underline{\tau}\right)\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right)^\alpha\right)^2\right]\right)^{1/2} \\ & \cdot \left(E^*\left[g_v\left(\underline{\check{X}}^*_{k_2}\right)^2\right]\right)^{1/2}, \end{aligned} \tag{5.78}$$

for some $\underline{\tau} = \underline{\check{X}}^*_{l-m+1} + \lambda\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right)$, $\lambda \in (0,1)$. Indeed, this expression is of similar form as the terms in the proof of Lemma 5.9. Thus the essential point is to show an adapted version of (5.43) for the present situation. Further expressions then can be handled analogously. Using the arguments as in the proof of Lemma 5.9, one obtains

$$\begin{aligned} & \left\|\left(\underline{\check{X}}^*_{k_1} - \underline{\check{X}}^*_{l-m+1}\right)^{\alpha_1+\alpha_2}\right\|_1 \\ \leq \; & \prod_{r=1}^m E^*\left|\left(\sum_{j=k_1+M(n)}^{l-m+M(n)} \hat{\psi}_j\left(U^*_{k_1+r-1-j} - U^*_{l+r-1-j}\right)\right)^{\alpha_{1,r}+\alpha_{2,r}}\right| \\ = \; & \prod_{r=1}^m E^*\left|\left(\sum_{j=0}^{l-m-k_1} \hat{\psi}_{j+k_1+M(n)}\left(U^*_{r-1-j-M(n)} - U^*_{l+r-1-j-k_1-M(n)}\right)\right)^{\alpha_{1,r}+\alpha_{2,r}}\right| \\ = \; & \mathcal{O}_P\left(\rho^{(k_1+M(n))(|\alpha_1+\alpha_2|)}\right), \end{aligned} \tag{5.79}$$

which holds by Assumption 7. Furthermore, $E^*\left[g_v\left(\underline{\check{X}}^*_{k_2}\right)^2\right] = \mathcal{O}_P(1)$ is assured by Assumption 7. For the non-bootstrap covariance in (5.77) one proceeds exactly in the same way. Altogether this leads to

$$\begin{aligned} &Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right) \\ &= \mathcal{O}_P\left(N(n)^{-1} + \rho^{k_1+M(n)}\right), \end{aligned} \tag{5.80}$$

for all $k_1 \notin T_n$ and $k_2 \in T_n$. Of course, if $k_1 \in T_n$ and $k_2 \notin T_n$ this result is directly adaptable.

Now we have all preliminaries at hand to turn to the remainder term R_4. Using the previous considerations and the summability of the autocovariances

$$\begin{aligned} R_4 &= \frac{b}{n-m+1}\sum_{k_1=1}^{l-m+1}\sum_{k_2=1}^{l-m+1}\Big(Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ &\qquad - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big) \\ &= \frac{b}{n-m+1}\sum_{k_1=M(n)+1}^{l-m+1}\sum_{k_2=M(n)+1}^{l-m+1}\Big(Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ &\qquad - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big) \\ &\quad + \frac{b}{n-m+1}\sum_{k_1=1}^{M(n)}\sum_{k_2=1}^{M(n)}\Big(Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ &\qquad - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big) \\ &\quad + \frac{b}{n-m+1}\sum_{k_1=1}^{M(n)}\sum_{k_2=M(n)+1}^{l-m+1}\Big(Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ &\qquad - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big) \\ &\quad + \frac{b}{n-m+1}\sum_{k_1=M(n)+1}^{l-m+1}\sum_{k_2=1}^{M(n)}\Big(Cov^*\left(g_u\left(\underline{\check{X}}^*_{k_1}\right), g_v\left(\underline{\check{X}}^*_{k_2}\right)\right) \\ &\qquad - Cov\left(g_u\left(\underline{\check{X}}_{k_1}\right), g_v\left(\underline{\check{X}}_{k_2}\right)\right)\Big) \\ &= \mathcal{O}_P\left(\frac{l}{N(n)} + \frac{b\,M(n)^2}{n} + M(n)\left(N(n)^{-1} + \rho^{M(n)}\right)\right) \\ &= \mathcal{O}_P\left(\frac{l}{N(n)} + \frac{b\,M(n)^2}{n} + \rho^{M(n)}\right), \end{aligned} \tag{5.81}$$

which converges to zero as n increases by the conditions on the rates ($l/N(n) \to 0$ and $\rho < 1$) and by the choice of $M(n)$. For the further remainder terms R_8 and R_{12}

one approaches in an analogue way. Thus, it follows altogether that

$$Cov^*\left(\frac{1}{\sqrt{n-m+1}}\sum_{t=1}^{n-m+1} g_u\left(\underline{\check{X}}_t^*\right), \frac{1}{\sqrt{n-m+1}}\sum_{s=1}^{n-m+1} g_v\left(\underline{\check{X}}_s^*\right)\right) \quad (5.82)$$
$$\to_{n\to\infty} \sum_{h=-\infty}^{\infty} Cov(g_u(\underline{X}_0), g_v(\underline{X}_h)) = (\Sigma_{q\times q})_{u,v} \quad i.p.$$

which concludes the proof of part *(i)*.

Now consider part *(ii)*, the central limit result. By Lemma 5.9 it suffices to show the result for

$$\frac{1}{\sqrt{n-m+1}}\sum_{t=1}^{n-m+1}\left(g\left(\underline{\check{X}}_t^*\right) - E^* g\left(\underline{\check{X}}_t^*\right)\right) \quad (5.83)$$

to conclude the proof. Define, for all $t = 1, \ldots, n-m+1$,

$$\underline{Z}_{n,t}^* := \frac{1}{\sqrt{n-m+1}}\left(g\left(\underline{\check{X}}_t^*\right) - E^* g\left(\underline{\check{X}}_t^*\right)\right) \quad (5.84)$$

which is a triangular array with $E^*\underline{Z}_{n,t}^* = 0$ for all $t = 1, \ldots, n-m+1$, and denote by $Z_{n,t,u}^*$ the u-th component of $\underline{Z}_{t,n}^*$, $u = 1, \ldots, q$. Due to the underlying block structure and the definition of $\check{X}_t^*$, the sequence is $2l$-dependent and does not yield stationarity. To proceed, a central limit result for $2l$-dependent triangular random variables is required where l is allowed to increase to infinity with sample size by some rate. In here we make use of Theorem 2.1 in Romano and Wolf (2000) and need to check their six conditions, say C1-C6. First, by using an straightforwardly adapted version of (5.81) compute

$$\begin{aligned} & B_{n,r,a}^2 \equiv B_{n,r,a}^2(u,v) \quad (5.85)\\ & := Cov^*\left(\sum_{t=a}^{a+r-1} Z_{n,t,u}^*, \sum_{s=a}^{a+r-1} Z_{n,s,v}^*\right) \\ & = Cov^*\left(\sum_{t=1}^{\lfloor \frac{r}{l}\rfloor l-m+1} Z_{n,t,u}^*, \sum_{s=1}^{\lfloor \frac{r}{l}\rfloor l-m+1} Z_{n,s,v}^*\right) + o_P(1) \\ & = \frac{\left\lfloor \frac{r}{l}\right\rfloor}{n-m+1}\sum_{h=-(l-m)}^{l-m}\left((l-m+1)-|h|\right) Cov\left(g_u\left(\underline{\check{X}}_0\right), g_v\left(\underline{\check{X}}_h\right)\right) + o_P(1) \end{aligned}$$

for all $u, v = 1, \ldots, q$, as before and note that

$$\left\lfloor \frac{r}{l}\right\rfloor l = \mathcal{O}(r). \quad (5.86)$$

Further compute

$$\begin{aligned} & B_n^2 \equiv B_{n,n,1}^2 \qquad (5.87)\\ & := Cov^* \left(\sum_{t=1}^{n} Z_{n,t,u}^*, \sum_{s=1}^{n} Z_{n,s,v}^* \right) \\ & = \frac{b}{n-m+1} \sum_{h=-(l-m)}^{l-m} ((l-m+1)-|h|)\ Cov\left(g_u\left(\underline{\check{X}}_0\right), g_v\left(\underline{\check{X}}_h\right)\right) + o_P(1) \end{aligned}$$

for all $u, v = 1, \ldots, q$, by the same manner. Now we check the required conditions for the central limit result.

Condition C1: $E^*|\underline{Z}_{n,t}^*|^{2+\delta} \leq \Delta_n$ for all $t = 1, \ldots, n-m+1$.
By Assumption 4 it holds $E|X_t|^{2+\delta} < \infty$. Then, as shown in the proof of part (i), the moments of X_t (and X_t^*) directly transfer to $g(\underline{X}_t)$ (and $g(\underline{X}_t^*)$) since the function $g(\cdot)$ yields Assumption 6 and Assumptions 7 and 8 are fulfilled. Thus, one obtains

$$E^*|\underline{Z}_{n,t}^*|^{2+\delta} \;=\; (n-m+1)^{-(2+\delta)/2} E^* \left| g\left(\underline{\check{X}}_t^*\right) - E^*\left(g\left(\underline{\check{X}}_t^*\right)\right)\right| \qquad (5.88)$$

Then choose $\Delta_n = C_Z \cdot n^{-(2+\delta)/2}$ for some suitable constant C_Z such as exemplarily $C_Z = q^2 \|\Sigma_{q\times q}\|_{\max}$.

Condition C2: $B_{n,r,a}^2\, r^{-(1+\gamma)} \leq K_n$ for all $a \in \{1, \ldots, n\}$ and for all $k \geq l$.
Let $\gamma = 0$ and recall equation (5.86). Then by the summability of the autocovariances and C_Z as before,

$$\begin{aligned} & B_{n,r,a}^2\, r^{-1} \qquad (5.89)\\ & \leq \frac{\mathcal{O}(1)\, l^{-1}}{n-m+1} \sum_{h=-(l-m)}^{l-m} ((l-m+1)-|h|)\ Cov^*\left(g_u\left(\underline{\check{X}}_0^*\right), g_v\left(\underline{\check{X}}_h^*\right)\right) \\ & = \mathcal{O}(n^{-1}l^{-1}) \sum_{h=-(l-m)}^{l-m} ((l-m+1)-|h|)\ Cov\left(g_u\left(\underline{\check{X}}_0\right), g_v\left(\underline{\check{X}}_h\right)\right) + o_P(1) \\ & \leq n^{-1} C_Z =: K_n \end{aligned}$$

Condition C3: $B_n^2\, (n\, l^\gamma)^{-1} \geq L_n$.
Use $\gamma = 0$ as before. Then we compute

$$\begin{aligned} B_n^2 n^{-1} \;&=\; \frac{1}{n} \sum_{h=-(l-m)}^{l-m} \left(\frac{(l-m+1)-|h|}{l} \right) Cov\left(g_u\left(\underline{\check{X}}_0\right), g_v\left(\underline{\check{X}}_h\right)\right) \qquad (5.90)\\ &\geq\; \frac{1}{2n}\, Cov\left(g_u\left(\underline{\check{X}}_0\right), g_v\left(\underline{\check{X}}_0\right)\right) =: L_n, \end{aligned}$$

Condition C4: $K_n L_n^{-1} = \mathcal{O}(1)$.
Using K_n and L_n as chosen in conditions C2 and C3 above, the present condition is straightforwardly fulfilled.

Condition C5: $\Delta_n \left(L_n\right)^{-(2+\delta)/2} = \mathcal{O}(1)$.
This condition immediately holds for Δ_n and L_n as defined in conditions C1 and C3.

Condition C6: $l^{1+(1-\gamma)(1+2/\delta)} n^{-1} \to 0$, as $n \to \infty$.
Since $\gamma = 0$ was chosen, it remains to consider $l^{2+2/\delta}\ n^{-1}$ which converges to zero by assumption.

Since all conditions of Theorem 2.1 in Romano and Wolf (2000) are fulfilled, the central limit theorem is applicable to the present scenario and leads to

$$B_n^{-1} \sum_{t=1}^{n-m+1} \underline{Z}_{n,t}^* \xrightarrow{\mathcal{D}^*} \mathcal{N}(0,1) \tag{5.91}$$

and furthermore, since $B_n^2 \to (\Sigma_{q\times q})_{u,v}$, for all $u, v = 1, \ldots, q$, one obtains

$$\frac{1}{\sqrt{n-m+1}} \sum_{t=1}^{n-m+1} \left(g\left(\underline{\check{X}}_t^*\right) - \theta^* \right) \xrightarrow{\mathcal{D}^*}_{n\to\infty} \mathcal{N}\left(0_q, \Sigma_{q\times q}\right). \tag{5.92}$$

By application of the delta technique, e.g. Billingsley (1979), the proof of part *(ii)* then is concluded for T_n^* with X_t^* replaced by $\check{X}_t^*$. Lemma 5.9 then gives immediately the result for T_n^* itself.

Part *(iii)* of the theorem is a direct consequence of part *(ii)* and Assumption 5. □

Acknowledgments

First and foremost, I want to express my deep gratitude to my supervisor Jens-Peter Kreiß for giving me the opportunity to write this thesis and for offering me a position in his group. He guided me thoroughly throughout this thesis, encouraged and supported me constantly in my research and far beyond.

Further, I like to thank Peter Brockwell and Efstathios Paparoditis for being trusting coauthors and for numerous very pleasant and fruitful discussions at different stages of this thesis. Efstathios Paparoditis further agreed to act as a referee for this thesis which I highly appreciate.

Many thanks also go to Anne Leucht and Alexander Lindner for the kind collaboration inside our institute and the pleasant work atmosphere.

I am very thankful to all of my actual and former colleagues for the remarkably nice atmosphere in our offices and the whole institute as well as for the time we spent together beyond.

Last but not least, I want to thank my parents and Frederike, besides my whole family and friends for the constant support during the last years. They encouraged and motivated me at any time it was necessary. Thank you all very much for that.

Bibliography

Akaike, H. (1969): Fitting autoregressive models for prediction. *The Annals of the Institute of Statistical Mathematics*, Vol. 21, pp. 243-247.

Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov and F. Csaki (eds.) *2nd International Symposium on Information Theory*, pp. 267-281. Akademiai Kiado, Budapest.

Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, pp. 716-723.

Anderson, T.W. (1994): *The Statistical Analysis of Time Series*. John Wiley and Sons, New York.

Andrews, D. (2004): The block block bootstrap. *Econometrica*, Vol. 72, pp. 673-700.

Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their applications in financial economics. *Journal of the Royal Statistical Society Series B*, Vol. 63, pp. 167-241.

Bartlett, M.S. (1946): On the theoretical specification and sampling properties of autocorrelated time series. *Supplement to the Journal of the Royal Statistical Society*, Vol. 8, pp. 27-41.

Bickel, P. and Freedman, D. (1981): Some asymptotic theory for the bootstrap. *The Annals of Statistics*, Vol. 9, pp. 1196-1217.

Billingsley, P. (1979): *Probability and Measure*. John Wiley and Sons, New York.

Bose, A. (1988): Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics*, Vol. 16, pp. 1709-1722.

Brockwell, P.J. (1994): A note on the embedding of discrete-time ARMA processes. *Journal of Time Series Analysis*, Vol.16, pp. 451-460.

Brockwell, P.J. (2001): *Continuous-time ARMA processes*. In: Handbook of Statistics, Vol. 19, pp. 249-276. Elsevier, Amsterdam.

Brockwell, P.J. (2009): Lévy-driven CARMA processes. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, Th. (Eds.), *Handbook of Financial Time Series*, pp. 457-480. Springer, Berlin.

Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods (2nd Ed.)*. Springer, New York.

Brockwell, P.J. and Lindner, A. (2009): Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Processs and their Applications*, Vol. 119, pp. 2660-2681.

Brockwell, P.J. and Lindner, A. (2012): Lévy-driven time series models for financial data. In: T. Subba Rao und C.R. Rao (Eds), *Handbook of Statistics*, Volume 30: Time Series Analysis: Methods and Applications, pp. 543-563. Elsevier, Amsterdam.

Brockwell, P.J. and Marquardt, T. (2005): Lévy-driven and fractionally integrated ARMA processes with continuous time parameter. *Statistica Sinica*, Vol. 15, pp. 477-494.

Brockwell, P.J. and Schlemm, E. (2011): Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. *Journal of Multivariate Analysis*, Vol. 115. pp. 217-251.

Brockwell, P.J., Davis, R.A. and Yang, Y. (2010): Estimation for non-negative Lévy-driven CARMA Processes. *Journal of Business and Economic Statistics*, Vol. 29, pp. 250-259.

Brockwell, P.J., Ferrazzano, V. and Klüppelberg, C. (2012): High-frequency sampling of a continuous-time ARMA process. *Journal of Time Series Analysis*, Vol. 33, pp. 152-160.

Brockwell, P.J., Kreiss, J.-P. and Niebuhr, T. (2014): Bootstrapping continuous-time autoregressive processes. *The Annals of the Institute of Statistical Mathematics*, Vol. 66, pp. 75-92.

Broersen, P.M.T. and de Waele, S. (2004): Finite sample properties of ARMA order selection. *Instrumentation and Measurement, IEEE Transactions*, Vol. 53, No. 3, pp. 645-651.

Bühlmann, P. (1994): Blockwise bootstrapped empirical processes for stationary sequences. *The Annals of Statistics*, Vol. 22, pp. 995-1012.

Bühlmann, P. (1997): Sieve bootstrap for time series. *Bernoulli*, Vol. 3, pp. 123-148.

Bühlmann, P. and Künsch, H.R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics*, Vol. 22, pp. 35-54.

Bühlmann, P. and Künsch, H.R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis*, Vol. 31, pp. 295-310.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistics from a stationary time series. *The Annals of Statistics*, Vol. 14, pp. 1171-1179.

Carlstein, E., Do, K.-A., Hall, P. Hesterberg, T. and Künsch, H.R. (1998). Matched-block bootstrap for dependent data. *Bernoulli*, Vol. 4, pp. 305-328.

Chan, K.S. and Tong, H. (1987): A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis*, Vol. 8, pp. 277-281.

Cohen, S. and Lindner, A. (2013): A central limit theorem for the sample autocorrelations of a Lévy driven continuous time moving average process with finite fourth moment. *Journal of Statistical Planning and Inference*, Vol. 143, pp. 1295-1306.

Dahlhaus, R. (1985). Asymptotic normality of spectral estimates. *Journal of Multivariate Analysis*, Vol. 16, pp. 412-431.

Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics*, Vol. 24, pp. 1934-1963.

Davis, W.W. (1977). Robust interval estimation of the innovation variance of an ARMA model. *The Annals of Statistics*, Vol. 5, pp. 700-708.

Davison, A.C. and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian and New Zealand Journal of Statistics*, Vol. 35, pp. 215-224.

Davison, A.C. and Hinkley, D.V. (1997): *Bootstrap methods and their application*. Cambridge University Press, Cambridge.

Doob, T.L. (1944): The elementary Gaussian processes. *The Annals of Mathematical Statistics*, Vol. 15, pp. 229-282.

Dowla, A., Paproditis, E. and Politis, D.N. (2003). Locally stationary processes and the local block bootstrap. In M. G. Akritas, and D. N. Politis (Eds.), *Recent advances and trends in nonparametric statistics* (pp. 437-444). North Holland: Elsevier.

Dunsmuir, W. and Robinson, P.M. (1981a): Estimation of time series models in the presence of missing data. *Journal of the American Statistical Association*, Vol. 76, No. 375, pp. 560-568.

Dunsmuir, W. and Robinson, P.M. (1981b): Parametric Estimators for Stationary Time Series with Missing Observations. *Advances in Applied Probability*, Vol. 13, pp. 129-146, 1981.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, Vol. 7, pp. 1-26.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. *Society for Industrial and Applied Mathematics*, Philadelphia.

Efron, B. and Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion). *Statistical Science*, Vol. 1, pp. 54-77.

Fan, J. and Yao, Q. (2003): *Nonlinear Time Series: Nonparametric and parametric Methods*. Springer, New York.

Fischer, W. and Lieb, I. (1988): *Funktionentheorie (5th Ed.)*. Vieweg, Wiesbaden.

Fowler, R.H. (1936): *Statistical Mechanics*. Cambridge University Press, Cambridge.

Francq, C. and Zakoïan, J.-M. (1998): Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference*, Vol. 68, pp. 145-165.

Francq, C. and Zakoïan, J.-M. (2009): Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis*, Vol. 30, No. 4, pp. 449-465.

Franke, J. and Härdle, W. (1992): On bootstrapping kernel spectral estimates. *The Annals of Statistics*, Vol. 20, pp. 121-145.

Freedman, D. (1984): On bootstrapping two-stage least squares estimates in stationary linear models. *The Annals of Statistics*, Vol. 12, pp. 827-842.

Götze, F., and Künsch, H. (1996): Second-order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics*, Vol. 24, pp. 1914-1933.

Hall, P. (1992): *The bootstrap and edgeworth expansion*. Springer Series in Statistics. Springer, New York.

Hall, P., Horowitz, J.L., and Jing, B.-Y. (1995): On blocking rules for the bootstrap with dependent data. *Biometrika*, Vol. 82, pp. 561-574.

Hannan, E.J. (1980): The estimation of the order of an ARMA process. *The Annals of Statistics*, Vol. 8, pp. 1071-1081.

He, S.W. and Wang, J.G. (1989): On embedding a discrete-parameter ARMA model in a continuous-parameter ARMA model. *Journal of Time Series Analysis*, Vol. 10, pp. 315-323.

Hull, J. and White, A. (1987): The pricing options on assets with stochastic volatilities. *The Journal of Finance*, Vol. 42, No. 2, pp. 281-300.

Huzii, M. (2001): Embedding a Gaussian discrete-time autoregressive moving average process in a Gaussian continuous-time autoregressive moving average process. *Journal of Time Series Analysis*, Vol. 28, pp. 498-520.

Jentsch, C. and Kreiss, J.-P. (2010): The multiple hybrid bootstrap – resampling multivariate linear processes. *Journal of Multivariate Analysis*, Vol. 101, pp. 2320-2345.

Jones, R.H. (1981): Fitting a continuous time model to discrete data. In *Applied Time Series Analysis II*, ed. D.F. Findley. Academic Press, New York, pp. 651-682.

Jones, R.H. (1985): Time Series Analysis with unequally spaced data. In *Time Series in the Time Domain*, Handbook of Statistics 5, ed. E.J. Hannan, P.R. Krishnaiah and M.M. Rao. North Holland, Amsterdam, pp. 157-178.

Kim, Y.-M. and Nordman, D.J. (2011): Large sample properties of a block bootstrap method under long range dependence. *Sankhya Series A - The Indian Journal of Statistics*, Vol. 73-A, pp. 79-109.

Kirch, C. and Politis, D.N. (2011): TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *The Annals of Statistics*, Vol. 39, No. 3, pp. 1427-1470.

Kokoszka, P. and Politis, D.N. (2011): Nonlinearity of ARCH and stochastic volatility models and Bartlett's formula. *Probability and Mathematical Statistics*, Vol. 31, pp. 300-312.

Kreiss, J.-P. (1988): *Asymptotical inference for a class of stochastic processes*. Habilitationsschrift, Universität Hamburg.

Kreiss, J.-P. (1997): *Asymptotical Properties of Residual Bootstrap for Autoregression*. Technical Report, TU Braunschweig. Preprint available on: https://www.tu-braunschweig.de/ stochastik/ team/ kreiss/ forschung/ publications.

Kreiss, J.-P. and Franke, J. (1992): Bootstrapping stationary autoregressive moving average models. *Journal of Time Series Analysis*, Vol. 13, pp. 297-319.

Kreiss, J.-P. und Neuhaus, G. (2006): *Einführung in die Zeitreihenanalyse*. Springer, Heidelberg.

Kreiss, J.-P. and Paparoditis, E. (2003): Autoregressive aided periodogram bootstrap for time series. *The Annals of Statistics*, Vol. 31, pp. 1923-1955.

Kreiss, J.-P., Paparoditis, E. and Politis, D.N. (2011): On the range of the validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, Vol. 39, No. 4, pp. 2103-2130.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, Vol. 17, pp. 1217-1241.

Lahiri, S.N. (1991). Second order optimality of stationary bootstrap. *Statistics and Probability Letters*, Vol. 11, pp. 335-341.

Lahiri, S.N. (1992). Edgeworth correction by moving block bootstrap for stationary and nonstationary data. In R. LePage, and L. Billard (Eds.), *Exploring the limits of bootstrap*. New York: Wiley.

Lahiri, S.N. (1993). On the moving block bootstrap under long range dependence. *Statistics and Probability Letters*, Vol. 18, pp. 405-413.

Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, Vol. 27, pp. 386-404.

Lahiri, S.N. (2003). *Resampling methods for dependent data*. New York: Springer.

Lahiri, S.N., Furukawa, K., and Lee, Y.-D. (2007). A nonparametric plug-in method for selecting the optimal block length. *Statistical Methodology*, Vol. 4, pp. 292-321.

Liang, G., Wilkes, D.M. and Cadzow, J.A. (1993). ARMA model order estimation based on the eigenvalues of the covariance matrix. *IEEE transactions on signal processing*, Vol. 41, pp. 3003-3009.

Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap*, eds. R. lePage and L. Billard. New York, Wiley, pp. 225-248.

Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, Vol. 10, No. 1, pp. 97-120.

Meyer, M. and Kreiss, J.-P. (2014): The Vector Autoregressive Sieve Bootstrap. To appear in *The Journal of Time Series Analysis*.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1994): Validity of blockwise bottstrap for empirical processes with stationary observations. *The Annals of Statistics*, Vol. 22, pp. 980-994.

Neumann, M. and Paparoditis, E. (2008): Simultaneous confidence bands in spectral density estimation. *Biometrika*, Vol. 95, No. 2, pp. 381-397.

Niebuhr, T. (2011): *Statistik für Prozesse mit strukturellen Beobachtungslücken*. Diploma thesis, TU Braunschweig, Braunschweig, 2011.

Niebuhr, T. and Kreiss, J.-P. (2014): Asymptotics for autocovariances and integrated periodograms for linear processes observed at lower frequencies. *The International Statistical Review*, Vol. 82, No. 1, pp. 123-140.

Niebuhr, T., Kreiss, J.-P. and Paparoditis, E. (2014): Autoregressive-aided block bootstrap. *In preparation.*

Nordman, D., Lahiri, S. and Fridley, B. (2007): Optimal block size for variance estimation by a spatial block bootstrap method. *The Indian Journal of Statistics*, Vol. 69, pp. 468-493.

Nordman, D. and Lahiri, S. (2014): Convergence rates of empirical block length selectors for block bootstrap. *Bernoulli*, Vol. 20, No. 2, pp. 958-978.

Paparoditis, E. (1996): Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis*, Vol. 57, pp. 277-296.

Paparoditis, E. and Politis, D.N. (2001): The tapered block bootstrap. *Biometrika*, Vol. 88, pp. 1105-1119.

Paparoditis, E. and Politis, D.N. (2002): The tapered block bootstrap for general statistics from stationary sequences. *Econometrics Journal*, Vol. 5, pp. 131-148.

Patton, A., Politis, D.N. and White, H. (2009): Correction to 'Automatic block length selection for the dependent bootstrap by D.N. Politis and H. White'. *Econometric Reviews*, Vol. 28, pp. 372-375.

Phillips, A.W. (1959): The estimation of parameters in systems of stochastic differential equations. *Biometrika*, Vol. 46, pp. 67-76.

Politis, D.N. (1993): On the maximum entropy problem with autocorrelations specified on a lattice. *IEEE Transactions on signal processing*, Vol. 41, No. 4, pp. 1715-1716.

Politis, D.N. and Romano, J.P. (1992): A circular block resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*, eds. R. lePage and L. Billard. New York, Wiley, pp. 263-270.

Politis, D.N., Romano, J.P. and Wolf, M. (1999): *Subsampling*. Springer, New York.

Politis, D.N. and White, H. (2004): Automatic block length selection for the dependent bootstrap. *Econometric Reviews*, Vol. 23, pp. 53-70.

Press, H. and Tukey, J.W. (1956): Power spectral methods of analysis and their application to problems in airplane dynamics. *Bell System Monographs 2606*.

Romano, J.P. and Thombs, L.A. (1996): Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 590-600.

Romano, J.P. and Wolf, M. (2000): A more general central limit theorem for m-dependent random variables with unbounded m. *Statistics and Probability Letters*, Vol. 47, pp. 115-124.

Singh, K. (1981): On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, Vol. 9, pp. 1187-1195.

Shao, Q.M. and Yu, H. (1993): Bootstrapping the sample means for stationary mixing sequences. *Stochastic Processes and their Applications*, Vol. 48, pp. 175-190.

Shao, X. and Wu, W.B. (2007): Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, Vol. 35, No. 4, pp. 1773-1801.

Shao, X. (2010): The dependent wild bootstrap. *Journal of the American Statistical Association*, Vol. 105, pp. 218-235.

Shibata, R. (1981): An optimal autoregressive spectral estimate. *The Annals of Statistics*, Vol. 9, pp. 300-306.

Tukey, J. (1958): Bias and confidence in not quite large samples (abstract). *The Annals of Mathematical Statistics*, Vol. 29, pp. 614.

Wu, C.F.J. (1986): Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, Vol. 14, No. 4, pp. 1261-1295.

Yang, Y. (2007): *Estimation for Lévy-driven CARMA processes*. Dissertation, Colorado State University, Fort Collins, Colorado.

Curriculum Vitae

Personal Data

Name	Tobias Niebuhr
born on	09/09/1987
in	Gifhorn

Professional experience and education

04/2011 - current	Research and teaching assistant at the Institute of Mathematical Stochastics, University of Technology Braunschweig
02/2011	Diploma in financial and business mathematics
10/2006 - 03/2011	Studies in financial and business mathematics at University of Technology Braunschweig
June 2006	Abitur (general university admission)
1993 - 2006	School

Lebenslauf

Persönliche Daten

Name	Tobias Niebuhr
geboren am	09. September 1987
in	Gifhorn

Werdegang

April 2011 - dato	Wissenschaftlicher Mitarbeiter am Institut für Mathematische Stochastik der Technischen Universität Braunschweig
Februar 2011	Diplom-Finanz- und Wirtschafts-Mathematiker
Oktober 2006 - März 2011	Studium der Finanz- und Wirtschafts-Mathematik an der Technischen Universität Braunschweig
Juni 2006	Abitur am Humboldt-Gymnasium zu Gifhorn
1993 - 2006	Schulausbildung